

Looking after number one

A merger of University College London and Imperial College, the top two research universities in Britain's capital city, may not in itself create a combined institution that is more internationally competitive.

The proposal to merge Imperial College and University College London (UCL), creating Britain's largest research university by far (see *Nature* **419**, 658; 2002), has been unveiled amid the warm and fuzzy hyperbole that tends to accompany such declarations of intent in business and in other fields. Much has been made of the world-beating prowess that such a merged institution could develop, of its potential ability to compete on a scale with the major US research universities, and of the vision of the principles involved.

After the smoke has cleared, however, it will be time for academics at both colleges to determine whether the merger of two of the country's top four research universities (the others are Oxford and Cambridge) will be in the best interests of their institutions, of London, or of British research as a whole.

A merger of Imperial and UCL would create a geographically bipolar institution in Britain's capital with around 28,000 students and an annual research income of some £400 million (US\$620 million) — almost as much as the largest US research universities, such as Johns Hopkins University in Baltimore and the University of Michigan. In terms of research clout it would tower above its closest competitors in Britain.

The architects of the merger proposal are two former industrial managers: Richard Sykes, rector of Imperial College and former chief executive of the drug company GlaxoWellcome, and Derek Roberts, provost of UCL and former deputy managing director of GEC, the erstwhile British electrical-engineering conglomerate. Both men have argued that the benefits of scale that would be generated by the merger would help the new institution to compete with world-class research universities elsewhere, particularly in the United States.

Spending power

However, the élite US research universities — Harvard, Yale, the Massachusetts Institute of Technology, Princeton, Stanford and the University of California, Berkeley, would rank on most people's lists — are not in fact the largest in terms of their annual research spending. In fact, few members of this group currently spend much more than either UCL or Imperial.

The largest research operators in the United States include the University of Michigan and the University of California, Los Angeles, which, while maintaining fine standards in both teaching and research, are seldom viewed abroad as the *crème de la crème* of the US system. And the pecking order of US universities, in terms of the funding they receive from an array of public and private sources, shows a broad equivalence between the top institutions. In 1999, the last year for which the National Science Foundation has compiled data, the University of Michigan, the largest institution, spent \$509 million; Harvard, the nineteenth-largest, spent \$326 million.

In other words, US policy-makers and university administrators have not found it expedient, either academically or politically, to allow one institution to assume a dominant position above the others. Yet rivals will fear that this is very much what Imperial and UCL have in mind.

The British government is expected to publish a discussion paper later this year on the future of the country's universities, and it will

probably encourage universities to consolidate their strengths on a regional basis. Mergers such as the one recently mooted between the University of Manchester and the University of Manchester Institute of Science and Technology would embody such consolidation (see *Nature* **416**, 114; 2002).

But it is not clear that a single major research institution for a city as demographically and economically dominant as London fits well within this framework. London is surely large enough to support a diverse academic life. The Manchester proposal is supposed to help the city and its hinterland support an internationally competitive university. But how can it even compete domestically if Imperial and UCL create a behemoth with four times as many resources as anyone else in the country except Oxford and Cambridge?

Open debate

Sykes and Roberts each have experiences that suggest that big is not always beautiful. Sykes was involved in the merger in December 2000 between GlaxoWellcome and SmithKline Beecham, one of a recent spate of drug-industry mega-mergers whose performance to date has fallen short of investors' hopes. The merged company's share value has fallen from £19 at the time of the merger to less than £14 this week. And Roberts' former employer, GEC, has effectively collapsed, after buying up and consolidating every UK defence-sector player it could get its hands on, including such once-esteemed operations as Plessey and Ferranti.

None of which is to suggest that the merger of the London institutions is without merit. It is probably true, for example, that Imperial's attractiveness to undergraduates is limited by its lack of involvement in the social sciences and humanities, which a merger with UCL would provide. The two institutions appear at first glance to be a reasonably good fit. They have done well to announce their intentions at an early stage, and must now allow their academics to participate in as open a debate as possible on the merits of the proposal.

The governing councils that will be asked to endorse the merger proposal before the end of the year — and perhaps also the British parliament, which must then legislate to enable it — must consider what the merger is likely to achieve for the institutions, their staff, their students and the wider community.

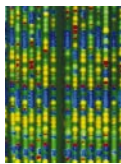
Both colleges' largest problems in competing internationally are an ageing infrastructure and a difficulty in generating fresh income beyond the narrow constraints of UK public funding. Merger proponents will argue that a single, dominant university in London will be better able to attract such funds from international sources. Much of the debate on the merger will focus on whether that is indeed the case.

UCL and Imperial College each have distinct identities that inspire strong loyalty among students, staff and alumni. The pressing need for reform and flexibility at UK universities was highlighted only last week by Imperial's statement that it plans to introduce hefty student tuition fees if the government gives it the go-ahead. But even in such a climate — and even if it is accepted that financial viability is the dominant consideration — these identities are of value, and should not be discarded lightly. ■





Scope for growth
Spain needs more
astronomers to meet
its project plans
p766



Legal battle
SuperScript enzyme
patent challenged
in court
p767



Equal rights
Japanese women
demand a level
playing field
p768



Fictional fellow
Sherlock Holmes
honoured by Royal
Society of Chemistry
p771

Universities face cash shortfall as stock-market slide hits charities

Declan Butler

When stock markets surged in the 1990s, philanthropic donations became one of the fastest-growing sources of funding for research and new laboratories throughout the world. But with stocks now staggering towards their third consecutive yearly loss, this growth has gone into reverse — sending shock waves through labs that have depended on it.

Universities and colleges in the United States received about one-sixth of the \$212 billion that Americans gave to charity in 2001, according to *Giving USA*, a report compiled by the Center on Philanthropy at Indiana University, Bloomington.

Although hard data for 2002 are not yet available, Vance Peterson, president of the Washington-based Council for Advancement and Support of Education, predicts that this year will be “the first year in the last 30 where giving towards higher education and research has dropped”. According to one survey, eight of the top ten US foundations reduced their grant-giving last year, and more cuts are expected.

Instead of planning for growth, most of the talk at philanthropic organizations is of toughing out the current crisis by making cuts or freezes in jobs, grants, building and research, until the economic outlook brightens. But the situation varies widely between organizations. The Seattle-based Bill and Melinda Gates Foundation, for example, which has pledged large sums to AIDS- and malaria-vaccine research, reports that most of its \$24-billion endowment is held in bonds, and has therefore been unaffected by the slump in share prices.

Others are less well placed. The biggest losers include foundations set up by a new breed of ‘Silicon Valley’ philanthropist (see *Nature* **410**, 140–143), whose wealth was based on high-technology stocks. The David and Lucile Packard Foundation in Los Altos, California, for example, has seen its assets — mostly Hewlett Packard stocks — fall from \$13.1 billion in 1999 to just \$3.5 billion. It intends to continue supporting activities it views as top priorities, including its Fellowships for Science and Engineering Program



Hard figures: falling stock values have cut into the resources of many philanthropic organizations.

and the Monterey Bay Aquarium Research Institute. But the foundation says that it is reducing its staff of 160 by half, and will distribute \$200 million in grants in 2003, down from \$250 million this year.

Some smaller foundations could be extinguished altogether. The San Jose-based Kirsch Foundation, for example, which was founded in 1999 by the Silicon Valley entrepreneur Steven Kirsch, has seen the value of its endowment fall from \$95 million in 2000 to around \$15 million. According to the foundation's president Kathleen Gwynn, its operating budget will exhaust these funds in a few years if markets do not recover and Kirsch is unable to provide more money.

The Howard Hughes Medical Institute — the largest US research philanthropic organization, with an annual budget of \$665 million — is currently “assessing the impact of sustained market volatility and lowered investment income on its programmes in biomedical research and science education”, according to spokeswoman Avice Meehan. It has frozen recruitment at its headquarters in Chevy Chase, Maryland. Other large US philanthropic agencies declined to comment, but the Rockefeller Foundation, for example, reduced its grant funding by one-fifth to \$120 million in 2001.

Across the Atlantic, the portfolio of Europe's biggest research charity, the London-based Wellcome Trust, fell from £15 billion (US\$23 billion) in 2000 to £12 billion last year, and may drop again this year.

Most charities try to even out their ups and downs by managing funds on a five-to-ten-year basis, says Diana Garnham, chief executive of the UK Association of Medical Research Charities. “There is unlikely to be a sudden drop in their research programmes, she says, “but if the stock market continues to underperform, then we can expect to see a downturn.” The impact will be most keenly felt by smaller charities, she adds.

Wealthy research universities face a double blow: the value of their endowment funds is falling in many cases, even as the incoming flow of new money dries up. US universities experienced an average drop in their endowments last year of 3.6%, according to figures compiled by the National Association of College and University Business Officers.

Fund managers had some success in limiting the damage, though — the Standard & Poor's (S&P) 500 stock index, the best measure of US stock values, fell by 14.8% in the same period. But with the S&P down by almost a quarter so far this year, university endowments are still shrinking in value. ■

B. MATTHEWS/AP

Fatal rocket accident puts question mark over Soyuz's safety

David Adam, London

Russian engineers are urgently trying to determine the cause of the accident that wrecked a Soyuz rocket shortly after take-off on 15 October, killing a soldier and destroying its payload, a European Space Agency (ESA) satellite.

A modified version of the same type of launcher is due to take three cosmonauts to the International Space Station (ISS). Russian space-agency officials have delayed the launch until 31 October to allow for their investigation.

Soyuz is viewed as a relatively cheap workhorse, and has been used to ferry people and equipment to the ISS about a dozen times since the beginning of last year. The unmanned version of the Soyuz launcher has a history of problems, although the more advanced manned rocket has an excellent safety record.

"No expense is spared when humans travel to space," says Georgy Poleshuk, deputy director of the Russian Space Agency.

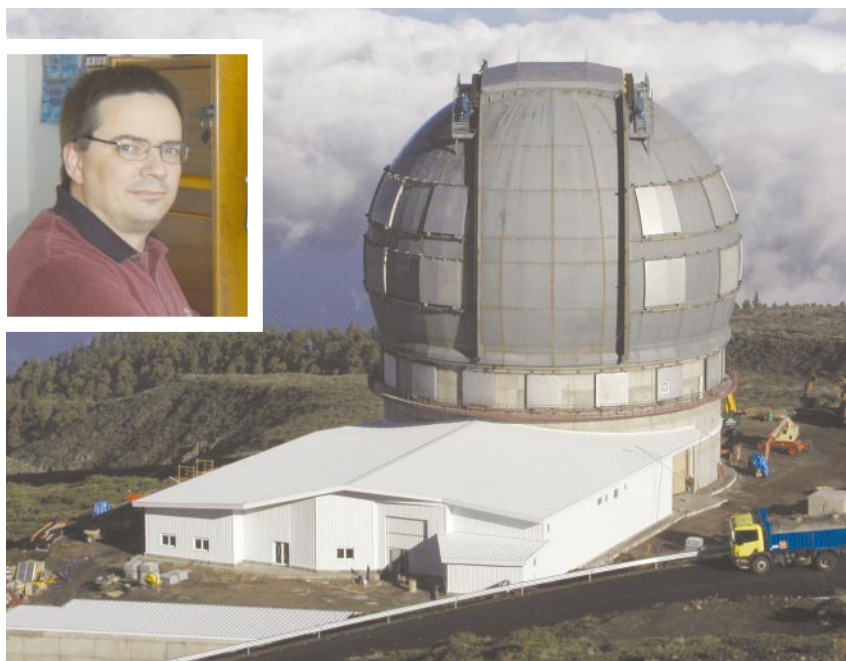
The accident, which occurred about 20 seconds after lift-off from the Plesetsk cosmodrome, 750 miles north of Moscow, showered the launch site and surrounding area with burning debris. The results of the accident investigation should be known this week. Early reports suggest that one of the rocket's first-stage engines failed to ignite.

The launcher was carrying Foton-M1, an ESA research satellite. "One of the advantages of this satellite was that some of the more delicate experiments could be placed on-board just a few hours before launch, so we had a number of researchers on-site at the time," says Dieter Isakeit, manager of ESA's ISS User Center in Noordwijk, the Netherlands.

All of the researchers were a safe distance away when the explosion took place, although a 20-year-old Russian soldier was killed and eight others were injured.

The satellite, which was designed for a 15-day mission, was carrying dozens of experiments covering a wide range of disciplines, including biology, crystal growth and exobiology.

Martin Miller, a space researcher at the Open University in Milton Keynes, UK, is part of a team that was aiming to investigate how meteorites burn up in the atmosphere. "Although in monetary terms our experiment was pretty small beer, this will cause significant delays and disruption to our research," he says. ■



Xavier Barcons is worried that Spain will not be able to make full use of the Gran Telescopio Canarias.

Spain's staff shortages leave astronomy plans up in the air

Monica Salomone, Madrid

Spanish astronomy departments are not hiring the people needed to take full advantage of the country's recent energetic involvement in international telescope projects, according to an extensive study undertaken for the government.

Spain is involved in a host of international astronomical projects, including the Atacama Large Millimeter Array (ALMA) under construction in Chile, the 10-metre Gran Telescopio Canarias (GTC) in the Canary Islands, set to begin operating in 2004, and several astronomical projects under the aegis of the European Space Agency (ESA).

But "the recruitment of new researchers is so low that we won't be able to profit from all this effort", says Xavier Barcons of the Spanish Council for Scientific Research, one of the authors of the study, which examined the performance of Spain's astronomers between 1999 and 2001 for the Ministry of Science and Technology.

Only half of Spanish astronomers hold permanent positions, the study has found, with the rest working under temporary contracts or fellowships. And in the period studied, the number of researchers increased by less than 2% per year. "Current human resources are scarce," the study concludes. "Without a substantial increase in the next years, we will be unable to face the challenges represented by the new instruments."

"Even in those areas in which Spanish

astronomy has a longer tradition, such as optical and infrared astronomy, the shortage is severe," says Barcons. "And soon we will have 90% of the time of the GTC. We are not going to be able to take advantage of the resources."

Barcons adds that although Spain pays for 7% of ESA's scientific programme, Spanish astronomers only get about 2% of the available observing time on the space agency's XMM-Newton X-ray telescope. Overall, Spain has 11.5 astronomers per million inhabitants, the study says, whereas Britain, Germany and France had 23, 17 and 16, respectively, in 1998.

The study's authors based their findings on an analysis of the scientific literature, data gathered from the 34 centres and university departments, and a questionnaire sent to 450 researchers, 40% of whom responded. In the period studied, each Spanish astronomer published an average of one paper per year in journals whose papers are cited elsewhere an average of just over three times each. The Spanish papers were cited almost five times each on average — indicating, the study's authors say, that Spanish astronomy is in good scientific shape.

Respondents to the questionnaire ranked the creation of more permanent posts for researchers as the top priority for the field, followed by more reliable research funding, and then full Spanish membership of the European Southern Observatory. ■

Super-enzyme patents get their day in court

David Cyranoski, Tokyo

A fierce patent battle over an enzyme widely used by biologists to create gene libraries reached its climax earlier this month.

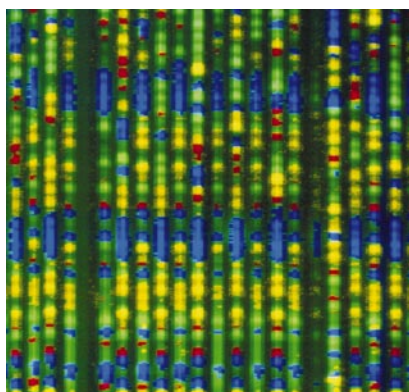
The case, which was held in Delaware's district court in Wilmington, concluded on 11 October. A ruling that will determine whether cheaper imitations of the popular SuperScript enzyme could become more readily available is expected within the next few months.

At present, biotechnology firm Invitrogen of Carlsbad, California, has four patents associated with SuperScript, and the company dominates distribution of the enzyme in the United States.

SuperScript is a reverse transcriptase used to turn molecules of messenger RNA (mRNA), taken from cells of plants and animals, into complementary DNAs (cDNAs). These cDNAs represent all or part of a gene, and are easier to work with in the lab than mRNA. By assembling the cDNAs into libraries, researchers can study how genes express themselves, for example, during different stages of development or disease (see *Nature* 419, 3–4; 2002).

But it is hard to ensure that all of the mRNA is transcribed — especially as reverse transcriptases tend to degrade mRNA while making cDNAs. SuperScript has been refined to minimize this degradation, making it a popular and effective tool.

The Delaware court case saw Invitrogen's patents challenged by BD Biosciences Clontech of Palo Alto, California. During the week-long hearing, Clontech argued that the



With complements: the SuperScript enzyme is used to generate cDNAs (above) for gene libraries.

patents are too narrow to cover SuperScript, that Invitrogen has misrepresented SuperScript features such as the degree to which it can reduce mRNA degradation, and that Invitrogen's protectionist strategy violates anti-trust laws. Clontech wants the judge to hit Invitrogen with massive fines — \$500 for each of the thousands of times the enzyme has been used — and to give half to Clontech as compensation.

Invitrogen officials reject all three claims but admit that the company is highly protective of SuperScript. "It is one of our core technologies," says a spokesman for the firm. Since 1996, Invitrogen has sued at least six companies, including Clontech, for reselling its reverse transcriptase or its products in violation of the purchase agreement. An

initial victory in this case by Clontech, which invalidated Invitrogen's patent coverage, was successfully appealed against and the dispute continues in a separate Maryland court case.

An unfavourable ruling for Invitrogen in Delaware could set a precedent for other pending cases, such as that between the company and Stratagene of La Jolla, California. It would also open up the US market to those who have stayed out in deference to Invitrogen's patents, observers of the case say.

In addition, a ruling against Invitrogen might free up the distribution of cDNAs and cDNA-based analytical tools. This could include microarrays made using SuperScript, some of which are only available to academic researchers because of the licensing agreements that Invitrogen demands from commercial collaborators.

Biotechnology firms, such as Tokyo-based Dnaform, which distributes the mouse cDNA libraries for Japan's Institute of Physical and Chemical Research (RIKEN), have complained that the terms of these agreements are too onerous. Toshizo Hayashi, Dnaform's president, claims that Invitrogen refused to grant it a licence to distribute the mouse cDNAs in the United States.

Hayashi says that no agreement has been reached and that the mouse cDNAs cannot be distributed to the United States unless an academic collaboration with RIKEN is established. But Alan Hammond, Invitrogen's patent lawyer, says that the company's protection of its patents is justified by its research efforts. "It's the cost of innovation," he says. ■

Malta provides loophole for breast-cancer screen

Erika Check, Washington

A Maltese biotechnology firm is set to challenge the potential monopoly on breast-cancer screening held by US company Myriad Genetics.

Synergene, which was set up in Malta in 2000, claims to have found a loophole in Myriad's patents, and says that it will launch a diagnostic test for breast cancer next January.

Myriad, which is based in Salt Lake City, Utah, owns patents on the genes *BRCA1* and *BRCA2*, and on tests to detect mutations in these genes. Such mutations are thought to account for a sizeable number of breast-cancer cases.

But Myriad has invoked the wrath of doctors and researchers across Europe by trying to stop them from using their own tests to search for *BRCA* mutations.

Synergene says that Myriad's patents do not apply in Malta because the country has

not signed the European Patent Convention, and Myriad has not applied for separate patents on its diagnostic tests there. Synergene will charge US\$2,000 for its new test, which will be similar to that offered by Myriad, and will be performed at Synergene's lab in Malta, says Patrick Willems, the firm's medical director.

Currently, doctors wanting to use Myriad's diagnostic tests must send samples to the company's laboratory in Salt Lake City. Alternatively, they can order similar analyses in Europe, as many labs have devised their own tests for mutations in the *BRCA* genes. But Willems says that he was frustrated by the European tests at his former job as a clinical geneticist at the University of Rotterdam. "We had to wait four months to get the results back," he says.

Willems adds that European labs may soon have to stop offering their home-

brewed tests. European researchers, patient groups and governments have lodged formal complaints against Myriad's patents (see *Nature* 413, 95–96; 2001). But if their effort fails, academic labs may have to pack up their test kits.

"We will be an alternative to Myriad, even if Myriad succeeds in enforcing its patent in Europe," Willems says.

But some opponents of Myriad's patents are unimpressed. Dominique Stoppa-Lyonnet, a doctor at the Curie Institute in Paris, says that Myriad might be able to prosecute French geneticists who buy tests from Malta. And French researchers probably won't want to pay Synergene's prices, she says, even though the company will charge \$760 less than Myriad per test. "It is an interesting situation, but I am afraid it is just another way to make a lot of money," says Stoppa-Lyonnet. ■

Bioprospectors turn their gaze to Canada

Rex Dalton, Kelowna

A Californian bioprospecting company that met resistance to its search for commercially valuable organisms in US national parks is now turning its attention to Canada.

Diversa of San Diego announced its latest plan at the World's Indigenous Peoples conference, held last week in Kelowna in British Columbia, Canada. The conference brought researchers, lawyers and tribal leaders together to discuss the status of intellectual property rights on traditional knowledge.

The company's proposal — believed to be the first of its type in Canada — is to study organisms at paper- or pulp-processing facilities on private land at an undisclosed location. But company officials say they would also like to search public land in Canada.

Government officials say they would welcome this as a starting point for wider exploration, which might lead to the commercial exploitation of either land or knowledge held by Canada's indigenous people.

In 1997, Diversa secured an agreement with the US National Park Service to start bioprospecting at Yellowstone National Park in Wyoming, triggering a still-unresolved dispute over this use of national parks. The Edmonds Institute, a small environmental group based in Washington state, sued the government, successfully calling for an environmental review of the deal, which is due to be completed next summer.



Dip in profits: Diversa's Eric Mathur collecting Yellowstone Park microbes before work stalled.

Diversa's plan to explore in Canada could set the stage for a similar conflict in a nation where suspicion of bioprospecting runs deep, especially among indigenous people. At the Kelowna conference, some speakers described the search for commercially useful natural organisms as "biocolonialism".

Beth Burrows, director of the Edmonds Institute, says she assumes that the Canadian government and public-interest groups "will do all they can to make sure the public is made aware of every step taken, every deal proposed and every agreement made — all in sufficient time for all the public to comment on how their commons are used".

Last month, Diversa officials wrote to the

government agency Environment Canada seeking authority to begin bioprospecting.

Jock Langford, senior policy adviser at Environment Canada, says the agency is pleased that Diversa contacted the government before starting work on private land. Government officials are reviewing the matter to see which agency has jurisdiction over lands selected by Diversa, he adds. Indigenous groups might also become involved through their long-standing claims to disputed land, some of their representatives say.

Leif Christoffersen, Diversa's biodiversity coordinator, says that there seems to be no law governing bioprospecting on private land in Canada, but that the status for public land is unclear. "We are interested in exploring public lands, but not until we establish a link with the ministry," says Christoffersen.

Diversa is seeking organisms that might form the basis of a viable biotechnology product. The *Taq* polymerase enzyme for the polymerase chain reaction, for instance, now used routinely by thousands of researchers, originated from heat-resistant bacteria found at a geyser in the Yellowstone Park.

Diversa has previously cast its net in developing countries, including Costa Rica, Indonesia, Kenya and South Africa. In Mexico, the company retreated after tribes objected to other bioprospectors. Diversa still holds samples from there in a freezer, but says that it is not using them for research. ■

M. MILSTEIN/AP

Marie Curie doesn't live here, Japanese women say

David Cyranoski, Tokyo

Japan's female researchers want respect and equal opportunity — without having to meet unreasonable standards of achievement.

That is the prevailing sentiment that arose from the landmark first meeting of a Japanese liaison council for gender equality on 7 October, which brought together 29 academic societies to find ways of improving the status of female researchers in Japan.

Education minister Atsuko Toyama — one of four women in the government's 18-member cabinet — opened the meeting, which members hope will mark the start of a serious effort to tackle a long-term problem. The percentage of women scientists tails off into single figures in the upper levels of both industrial labs and universities.

Speakers outlined various ongoing efforts to create a more even playing field, including the introduction of crèches at society meetings and appeals to government, universities and granting agencies to create more transparent evaluation and recruitment systems.

But the meeting heated up when several delegates objected to a mission statement honouring cancer researcher Marie Curie for "having proven the outstanding ability of women in the natural sciences". "The statement implies that only special women with tremendous abilities can compete with men," said Kay Domen, a researcher on blue semiconductor lasers at Fujitsu Quantum Devices. "Regular women also need to be able to compete." It is now being revised.

The committee's first task is to conduct a nationwide survey. One of the organizers, Kazuo Kitahara, president of the Physical Society of Japan, says she is confident that the committee will get a grant from the education ministry to collect statistics on the number of women in different fields, their success in obtaining grants, and even information about the prevalence of sexual harassment. Female researchers say that a lack of detailed information has exacerbated their problems (see *Nature* 410, 404–406).

Despite some scepticism from meeting participants about its chances of making a

difference, the committee has at least raised awareness. For example, the Society of Chemical Engineers, Japan, now recognizes that only 2% of its members are female — and none of these women holds a senior position in industry or academia. "Until now we had never thought about building a society based on equal participation by men and women," the society admitted in a statement. ■



Reaching for the prize: but high positions still elude most women scientists in Japan.

T. MATSUMOTO/AP

Medical funding group calls for clamp-down on hype

David Adam, London

Researchers who talk to the press prematurely about unpublished research could soon face harsher sanctions than the odd disapproving glance from a colleague. Under research misconduct guidelines just released by an association of British medical charities, they could be blacklisted for funding, the head of the association says.

Diana Garnham, chief executive of the London-based Association of Medical Research Charities (AMRC), which issued the guidelines on 17 October, says its members are fed up dealing with the fallout from over-hyped or misleading results. "Scientists don't do their work in a vacuum," she says. "There is an audience for whom their work is directly relevant, and they need to bear that in mind."

The guidelines state simply that researchers should be "especially careful" when discussing incomplete work, and are intended to coax universities and other research institutions into drawing up their own rules. They also point out that the aim of disseminating research "should not be primarily to seek publicity for the researcher, the research institution or the funder". From January next year, the AMRC says it will recommend that its members fund only researchers at institutions that have published specific standards for sound scientific conduct.

The AMRC counts major research funders such as the Wellcome Trust and Cancer Research UK among its members — but it remains unclear how much impact the new proposals will have. The Wellcome Trust has already issued guidelines of its own, which came into force earlier this month (see *Nature* **412**, 667; 2001).

Robert Terry, a senior policy adviser at the trust, says that the charity is unlikely to act immediately on the new recommendations.

It seems to be an unfinished threatening letter...



Research on the health risks of the MMR vaccine reached the press before it was finalized.

"They are a useful addition but I don't think we would go that far," he says. The existing peer-review process of grant applications can already take into account bad publication practices or the over-promotion of results, Terry claims.

But the AMRC says that more action is needed to deal with what it says are a small but nonetheless significant number of incomplete research findings — such as those concerning a possible link between childhood autism and the combined measles, mumps and rubella (MMR) vaccine — that are promoted by researchers and then heavily reported in the press.

"When scientists want to protect the commercial outcome of their research they stay in control of the timing of its release, so why should this be any different?" says Garnham. "We don't go as far as saying that their funding should be cut off — but I think some of our charities would be willing to do that."

Others observers say that it will be hard to hold researchers accountable for premature release of their results, still less for their misinterpretation by the media. "Some sort of quality-control mechanism is needed," says Bob Ward, a spokesman for the Royal Society, which will shortly announce its own inquiry into the dissemination of research results, "but it's not a straightforward issue."

♦ www.amrc.org.uk

National academies slam Bush proposal for data security

Erika Check, Washington

America's scientific elite has issued a stern warning about what it says are the pitfalls of the Bush administration's proposal to create a new category of sensitive, but unclassified, technical information.

The presidents of the National Academy of Sciences, National Academy of Engineering and the Institute of Medicine released a joint statement on 18 October rejecting the proposed category, which the White House says is needed to prevent the misuse of certain scientific and technical information. The administration hasn't released details of what kinds of information would be included in the category, which it calls 'sensitive homeland-security information' (see *Nature* **418**, 906; 2002).

"Experience shows that vague criteria of this kind generate deep uncertainties among both scientists and officials responsible for enforcing regulations," the academies' statement says. "The inevitable effect is to stifle scientific creativity and to weaken national security."

At a hearing of the House Committee on Science on 10 October, John Marburger, director of the White House Office of Science and Technology Policy (OSTP), sought to assuage scientists' concerns about the proposal. He said that the category guidelines, when released, would not cover results from basic research. The designation is for the type of information held by the government that is not routinely released to the public, such as law-enforcement data, he told the hearing.

But Bill Colglazier, executive officer of the National Academy of Sciences, says that without written guidelines, scientists can't accept Marburger's assurances. "The concern is that if you leave this category very vague and amorphous, it could end up causing problems in the research community," says Colglazier.

Biologist M. R. C. Greenwood, the chancellor of the University of California, Santa Cruz, agrees that it would be a bad idea to set up a category of unclassified but restricted information. "We would be better off building up science and technical enterprise so we have talent and ideas to outwit any threat that comes this way, as opposed to trying to keep information from escaping," says Greenwood, who served as an associate director of the OSTP in the Clinton administration.

♦ www.nas.edu

Senate move starts run up for triple jump in NIH budget

Washington With the doubling of the budget for the National Institutes of Health (NIH) not yet complete, some in Congress are already talking about giving the agency an even bigger boost.

On 17 October, Senator Arlen Specter (Republican, Pennsylvania) introduced a 'Sense of the Senate' resolution — a non-binding measure that senators can use to pledge support for an issue — that asks his colleagues to boost the NIH's funding to \$41 billion by 2008, tripling the agency's budget since the fiscal year 1998.

The NIH's backers outside Congress appreciate Specter's support, but say that their current priority is trying to get this year's budget passed. Many are beginning to worry that the NIH will not get the \$3.7-billion increase it needs to complete a doubling of its budget in the fiscal year 2003.

US monitor of clinical trials to step down next month

Washington Greg Koski, head of the US Office for Human Research Protections, has said he plans to quit his post at the end of November.

The office was created in 2000 to improve federal monitoring of clinical trials, and Koski is its first director. Under his leadership, the agency has started developing conflict-of-interest guidelines and has issued guidance for groups that cannot give consent to research, such as prisoners and children.

Koski's announcement, which was made on 9 October, follows the Bush administration's decision to wind up the National Human Research Protections Advisory Committee, a federal advisory panel. Some observers have suggested that the two events may be linked, but others pointed out that Koski had promised changes but was not able to implement them.



Time to go: Greg Koski returns to Harvard.

"There's been a sense of a lot of bombast but not a lot of follow-through," said one federal official who asked not to be named.

Koski will now return to his former post as an associate professor at Harvard Medical School and an anaesthesiologist at Massachusetts General Hospital in Boston.

Cancer audit reveals spending inequalities

London Just 2% of British cancer-research funding is spent on preventing future cases of the disease, the first major audit of the sector has revealed.

The London-based National Cancer Research Institute, an umbrella body for the government and the major charitable bodies that fund cancer research, revealed that 40% of the £257 million (US\$398 million) spent on cancer research each year goes on basic biology, 22% on treatment research and 16% on exploring possible causes. The finding that relatively low sums are spent on prevention, patient care and survival research prompted the institute to set up groups to see what further research could be done in these areas.

Despite the fact that lung cancer accounts for nearly a quarter of cancer-related deaths, research on the disease receives just 3% of available funds. Research into leukaemia, which is much rarer, takes nearly a fifth of the money.

Fast-track plan to double NSF funding hits the buffers

Washington A proposal to double the budget of the US National Science Foundation (NSF) over the next five years has temporarily stalled.

Bills to double the agency's money were introduced in the Senate and the House of Representatives earlier this year (see *Nature* 417, 209; 2002) by a bipartisan group advocating rapid growth for the agency. The Senate bill was undergoing a fast-track process, which requires unanimous approval, but Jon Kyl (Republican, Arizona) blocked its passage last week. Kyl is believed to be acting on behalf of the president's Office of Management and Budget, which wants a more modest three-year bill.

The bills' backers say that they will now try to bypass the block and bring the bill up for a vote when the Senate comes back into session in early November.

China's space visitors denied at final frontier

Washington Organizers of the World Space Congress, held earlier this month in Houston, Texas, have protested about the treatment of dozens of Chinese scientists and engineers, who were prevented from attending the meeting because of visa complications.

All but 2 of the 70 Chinese delegates were denied visas by the US Department of State on the grounds that they applied too late. The heads of the congress's two main sponsoring organizations, Marcio Barbosa of the International Astronautical Federation

and Gerhard Haerendel of the Committee on Space Research, wrote to National Academy of Sciences president Bruce Alberts last week to complain about the incident. Delegates from most countries had no difficulty with late applications, according to Barbosa's executive assistant, Oskar Klingl.

Klingl says that Chinese delegates and conference organizers only learned of the decision at the last minute, forcing the sudden withdrawal of most of the 40 Chinese papers that were to have been presented at the meeting.

Sherlock Holmes and the case of the chemistry fellow

London British chemists took an elementary step towards polishing their public image last week by making fictional detective Sherlock Holmes an honorary fellow of the Royal Society of Chemistry.

The society's own Dr John Watson presented a specially struck silver medal to the statue of Holmes that stands near Baker Street, London, the street on which Holmes lived. Also present was a Mastiff crossbreed hound, intended as "a reminder of the dog that haunted the Baskerville family" — a case that Holmes solved 100 years ago.

Holmes was not around to solve the mystery of why the society chose to honour



Elementary, my dear Watson: Sherlock Holmes statue is presented with the Royal Society of Chemistry's fellowship award.

him, so the society's chief executive David Giachardi offered an explanation. "Our particular interest is his love of chemistry and the way that he wielded such knowledge for the public good," Giachardi said.

Correction The News story on gene-therapy trials in severe combined immunodeficiency disease (*Nature* 419, 660; 2002) incorrectly stated that the US Food and Drug Administration (FDA) had decided that the trials should continue. In fact, an advisory committee to the FDA recommended that the trials be permitted to resume, but the FDA has not yet lifted their suspension.



Publish, and be damned...

Recent controversies over scientific fraud and other disputed findings have raised questions over the way in which journals select papers for publication. Is there a problem? And what more could be done to weed out dubious results? David Adam and Jonathan Knight investigate.

You browse through the latest issue of a journal and find a paper describing work from a competing group that you know to be riddled with holes. Your hackles begin to rise. Were the referees asleep? What was the editor thinking of?

Sometimes, it's only with hindsight that such feelings kick in. When a prominent researcher is accused of fabricating data, for instance, you might look back over the contested publications and see warning signs in almost every paper. In retrospect, those data really were too good to be true. So why did no one question their veracity when the papers were being reviewed?

Over the past few months, a series of high-profile controversies has brought such questions to the fore, throwing a spotlight on the workings of the journals that published the contentious work. Competition between scientists can tempt some individuals to conduct 'quick and dirty' experiments, rather than doing the job properly, in the hope of being the first to unveil startling new data.

In extreme cases, less scrupulous researchers may commit outright fraud. But are leading journals exacerbating the problem by competing to rush 'sexy' findings into print?

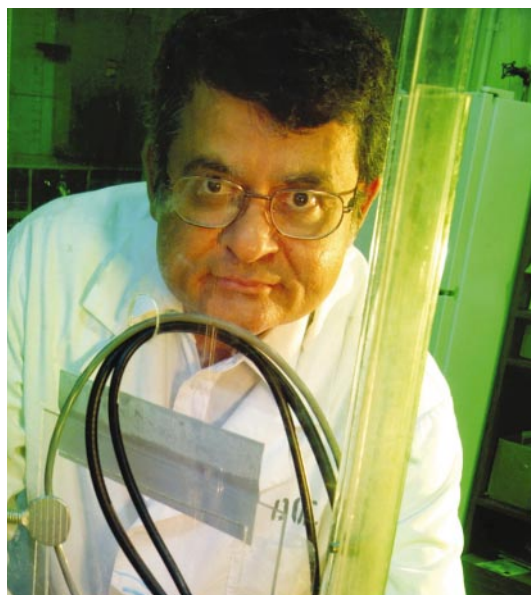
Accusations began to fly in March, when *Science* published a report¹ from scientists led by Rusi Taleyarkhan at the Oak Ridge National Laboratory in Tennessee who claimed to have triggered nuclear fusion in a beaker of organic solvent. The paper appeared to howl of protest, both from leading physicists who were sure that the authors were mistaken and from other researchers at Oak Ridge who had examined the work and claimed to have uncovered serious flaws.

A month later, *Nature* printed a brief statement² effectively disowning a paper³ it had published the previous year, which suggested that DNA from genetically modified maize had invaded the genomes of native Mexican varieties of the crop. The original paper, by David Quist and Ignacio Chapela of the University of California, Berkeley, provoked a political storm in Mexico. But

after publication, other experts argued that the findings were probably experimental artefacts.

In those two cases, researchers are arguing over whether papers' conclusions are justified by the data they contain — there is no suggestion of any misconduct. But it is the scandal surrounding the work of Jan Hendrik Schön of Bell Laboratories in Murray Hill, New Jersey, that has really set tongues wagging. Schön's research on molecular-scale electronic devices and induced superconductivity in carbon 'buckyballs' led to an avalanche of stunning papers — many in leading journals including *Nature* and *Science*. But we now know that he was the perpetrator of the biggest fraud ever to taint the physical sciences, fabricating and misrepresenting data on a massive scale⁴. And some researchers argue that the journals must shoulder some of the blame, for failing to scrutinize more closely the extraordinary claims coming from Schön's lab.

Each of these controversies has its partic-



L. FREEMAN/DOE

Too hot to handle? The conclusions of papers on nuclear fusion by Rusi Taleyarkhan (above) and on transgenic maize by David Quist (above, far right) and Ignacio Chapela have been challenged. Meanwhile, Jan Hendrik Schön (below) has been found guilty of fabricating data in several papers.

ular set of circumstances. But collectively, they are causing scientists and interested observers to ask whether erroneous or downright fraudulent papers are becoming more likely to be published; and whether editors and referees are doing enough to prevent the pollution of the scientific literature. After the Schön verdict, for instance, an article in *The Wall Street Journal* alleged that *Science* and *Nature* “are locked in such fierce competition for prestige and publicity that they may be cutting corners to get ‘hot’ papers”.

These are tough issues to address, because hard facts are difficult to come by. The US Office of Research Integrity did find more cases of misconduct in the biomedical sciences in 2001 than at any time since monitoring began in 1997, but this may simply be down to increased vigilance. And the frequency with which flawed, rather than fraudulent, papers are entering the literature is almost impossible to quantify. Certainly, few become a matter of public record. In the year to the end of September 2002, for example, *Nature* published just one retraction and two corrections that significantly altered a paper's conclusions.

Editors of leading journals reject the suggestion that standards are slipping in the face of heightened competition. “*Nature* has nothing to gain by the pursuit of glamour at the expense of scientific quality, considering, not least, the criticisms, corrections and retractions we would then habitually be forced to publish,” argued this journal in an editorial comment earlier this month⁵. And many researchers who were interviewed for



this article agree that the system is still working tolerably. If it ain't broke, they say, don't try to fix it.

But in the wake of the Schön scandal, critics of the status quo are speaking out. One of the most vocal is Nobel physics laureate Robert Laughlin of Princeton University in New Jersey. “In this case the editors are definitely culpable,” claims Laughlin. “They chose reviewers they knew would be positive.” Laughlin suspects that it was because of his open criticism of Schön and his co-authors, for failing to provide sufficient information about their methods to others who wanted to replicate the work, that he was not asked to review any of the papers. “It was well known that I was angry at these guys for not allowing their experiments to be reproduced,” Laughlin says.

Due process

Such charges are difficult to prove or disprove, given the confidentiality of the review process. But Karl Ziemelis, chief physical sciences editor at *Nature*, denies that referees were cherry-picked to ensure Schön's papers safe passage. “I can absolutely guarantee that we did not choose reviewers on the basis that we knew they would be positive,” he says. *Science*'s editor-in-chief Donald Kennedy has also rejected accusations that the Schön case reveals any shortcoming in the review process^{6,7}. But candidly, Ziemelis admits that *Nature*'s editors — like many physicists — were enthralled by Schön's work. “Given the exciting nature of the claims made by the papers, we were certainly hoping that the outcome would be positive,” Ziemelis says. But that is not unusual, he says, and tough refereeing ensures that “we are often disappointed”.

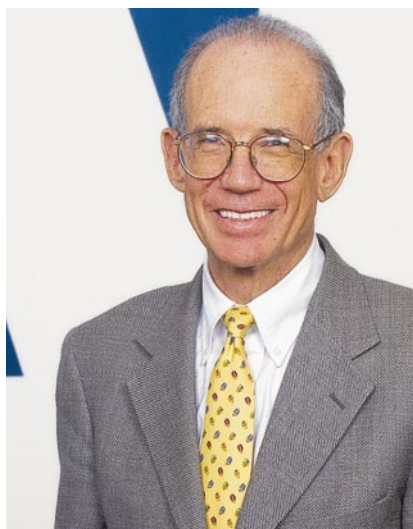
The success rate of Schön's submissions to *Nature* was “far from 100%”, Ziemelis adds. Of those that were eventually pub-

lished, he says, the referees delivered fundamentally positive reports, although they did take issue with some aspects of the papers — which were revised accordingly. But given that *Nature*'s editors were enthused by the claims coming from Schön's lab, would they ever have been willing to gloss over negative reviews and publish a paper anyway?

“It depends entirely on the basis of the negative comments,” says Ziemelis. There are essentially two types: technical and editorial. Ziemelis says that referees pointing out technical flaws that undermine the work will sink a paper — unless other reviewers whose expertise is sought on this specific point disagree. But when it comes to deciding whether a paper is interesting or important enough for *Nature*, editors might overrule a reviewer's negative comments. “The distinction between editorial and technical decisions is an important one, but it is often misunderstood,” Ziemelis says. “On editorial matters we have overruled



BELL LABS



Donald Kennedy denies that recent controversies reveal problems with the peer-review system.

both negative and positive referees.”

But the controversy surrounding Tale-yarkhan’s fusion paper blew up because of accusations that *Science*’s editors had overruled referees’ technical criticisms. Four of the referees asked by the journal to review the paper have taken the highly unusual step of speaking publicly about the confidential reports they supplied. “My problem with the paper was clearly technical — that the authors failed to provide sufficient evidence for what they were claiming,” says William Moss, a physicist at the Lawrence Livermore National Laboratory in California, who reviewed two different versions of the paper.

Another of the referees, Seth Putterman of the University of California, Los Angeles, goes further. “The earlier version of the paper had information that I claimed disproved their results and I pointed that out,” he says. But he claims that the offending data were simply removed from the final version.

Putterman and two other referees — physicist Larry Crum of the University of Washington in Seattle and chemist Ken Suslick of the University of Illinois at Urbana-Champaign — have even released their own critique of the work on the arXiv physics preprint server⁸. “Somewhere out there is a positive report from someone,” Putterman says. “*Science* should publish that report because then we’ll see what kind of information they went on to overrule four negative reviewers.”

That isn’t going to happen, says Kennedy. “We maintain our end of the confidentiality bargain about peer review,” he says, “so I can’t discuss the process specifically, except to say that the positive reviews outweighed the negative ones. Why else would we publish the paper?” Critics suggest that the kudos to be gained by the journal if the paper’s findings had been shown to be valid might have been a factor in the decision — a view that draws a forthright denial from Kennedy: “We at



Karl Ziemelis rejects claims that referees were chosen to ensure safe passage for Schön’s papers.

Science emphatically reject such charges.”

The top journals, rarely reticent in publicizing their latest hot paper or soaring impact factor, make easy targets for criticism when a paper subsequently comes under attack. But editors who have worked for these journals argue that scientists must share responsibility for any problems that exist. “There is sometimes pressure on referees to be quick rather than thorough,” says Philip Ball, a London-based science writer and a former physical sciences editor for *Nature*. “That’s an issue for the journals but also for the scientific community in general,” he says, noting that some authors try to play one journal off against another to get their papers published as quickly as possible.

Despite such concerns, standards of edit-

ing and refereeing are generally agreed to be higher at the elevated end of the scientific-publishing food chain. Further down the scale, fewer questions are asked. “There is an awful lot of literature pollution,” says Carole-Beth Stewart, an evolutionary geneticist at the State University of New York in Albany. “If people get rejected at *Nature*, they just work their way down the ladder, often ignoring all the reviewers’ comments.”

Damage limitation

Ultimately, it is to everyone’s advantage to keep ‘bad’ results out of the literature. Both scientists and journals rely on their reputations, and no one wants to have to retract a paper. So, in the light of this year’s controversies, could the system be tweaked to protect all involved from such embarrassment?

Journals manage their editorial processes in subtly different ways — with one important issue being whether they employ professional editors, or get working scientists to fulfil the role (see “Who should sit in the editor’s chair?”, below). But journals rely heavily for quality control on the efforts of the specialist referees asked to review individual papers. They are “unsung heroes”, says Nicholas Cozzarelli, editor-in-chief of *Proceedings of the National Academy of Sciences*.

Most scientists agree that the feedback from a careful referee is invaluable, regardless of the final decision on publication. “Many scientists don’t have the ability to step back from their own work,” says Anne Weil, an anthropologist at Duke University in Durham, North Carolina. But the refereeing process is meant also to spot crucial flaws in a paper that might affect its conclusions.

That apparently failed to happen with

Who should sit in the editor’s chair?

One topic of regular debate in scientific publishing is whether the job of editor is best done by a full-time professional or by a working scientist. Journals such as *Nature* and *Cell* employ a large number of editorial staff, usually lured from top research labs. These editors have considerable power — papers that fail to pass their initial screen don’t make it to peer review. But at many other journals, including most of those run by scientific societies, the editing role is handed to active researchers. *Science*, meanwhile, operates a hybrid system, employing a cadre of professional editors but also having papers screened by a board of reviewing editors — working scientists who have a major say in whether papers are sent out for review.

Full-time editors argue that they are not only able to devote their full attention to manuscripts, but are also more readily able to share information about the strengths and weaknesses of the reviewers they ask for advice. What’s more, full-time editors claim also to be unencumbered by professional bias. Working researchers may find themselves facing decisions that have ramifications for their own careers, such as whether to publish data that challenge a hypothesis that they have championed. And although it’s likely that most scientist-editors take the high road, professional editors never face the dilemma in the first place.

In response, proponents of editors as active researchers argue that their continuing hands-on expertise better equips them to decide whether a paper represents a significant advance in a given field. “We are proud of the fact that we don’t use professional editors,” says Nicholas Cozzarelli, a biochemist at the University of California, Berkeley, who serves as editor-in-chief of *Proceedings of the National Academy of Sciences*.

But no one has produced any solid evidence that one system is inherently superior to the other. Ultimately, any journal that applies rigorous editorial standards will continue to attract high-quality submissions. And with the top tier of journals including examples of both professional and working-scientist editorship, the debate shows no sign of reaching a definite conclusion.

Quist and Chapela's paper. The Berkeley researchers used two techniques to look for transgenic contamination in Mexican varieties of maize. The first was the standard polymerase chain reaction (PCR), a method for amplifying and detecting tiny quantities of a specific DNA sequence. The second, a variant called inverse PCR (i-PCR), was used to examine the DNA flanking these sequences to reveal their location in the genome. And it was the authors' claim that i-PCR had shown the transgenes to have fragmented and scattered throughout the genomes of the native maize varieties that caused such consternation. Groups opposed to agricultural biotechnology seized on the result, whereas scientists who are supportive of the technology began poring over the results in search of flaws.

Before long, the doubters found a problem. Close examination of the sequences amplified by i-PCR, deposited in the GenBank database, suggested that some were probably not associated with transgenes after all^{9,10}. Quist and Chapela then produced other evidence to support their original claim¹¹. But when this failed to convince one of the referees consulted by *Nature*, the journal published the exchange with its controversial note concluding that "the evidence available is not sufficient to justify the publication of the original paper"¹².

Crucial details

Johannes Fütterer, a plant scientist at the Swiss Federal Institute of Technology in Zurich, was a co-author of one of the critiques pointing out the problem with the sequences amplified by i-PCR, and claims that *Nature's* referees were remiss in letting the original paper through. "In this case, the sequences were the only piece of data that could prove what the authors claimed and they should have been checked," he says.

Ritu Dhand, *Nature's* chief biology editor, is reluctant to criticize the referees. "It was an unfortunate incident that bypassed us all," she says. Maybe so, but given the political sensitivity of the work, an awful lot of trouble and embarrassment would have been saved if at least one of the referees had followed the path that Fütterer suggests.

Quist and Chapela's paper was bound to become a focus for the contentious debate over genetically modified crops. Less obvious to those outside the Berkeley campus was the fact that the pair had campaigned against a multimillion-dollar deal between the university and the Swiss-based multinational Novartis — subsequently inherited by the spin-off agribiotech firm Syngenta — that gave the company privileged rights to exploit Berkeley's plant-science research. And the authors of the two published critiques of Quist and Chapela's i-PCR analysis included current or former Berkeley colleagues who backed the Novartis deal. In retrospect, given these circumstances, the paper was always likely to be combed for any flaws that

You can't refuse to publish something just because it doesn't feel right.

David Lindley

the referees had missed.

Whatever checks are applied, some papers are published when the referees, editors and even the authors suspect problems. David Lindley, a science writer in Virginia who previously worked as an editor for both *Nature* and *Science*, recalls one such *Nature* paper in 1991 — the first discovery of a planet beyond our Solar System¹². The planet appeared to complete a full orbit around a pulsar in exactly half the time taken by Earth to orbit the Sun. "The authors said: 'We know this is odd, but we can't find an obvious mistake.' One of the referees agreed there was something fishy," Lindley says. "But you can't refuse to publish something just because it doesn't feel right." Six months later, the red-faced authors discovered an error in their calculations. There was no planet, and the paper was retracted¹³.

In this case, Lindley says that it was impossible for him or the referees to discover what was an honest mistake. But what if authors are less than honest? Many editors have encountered papers that look 'too good'. Gene Wells, editor of *Physical Review Letters* for two decades until his retirement in 2001, recalls one case in which the referees all noted that the data seemed unnaturally free of noise. The manuscript was rejected, and Wells never heard about the paper again.

Wells' referees were alert to the possibility of data manipulation. But many scientists argue that referees cannot and should not be expected to view each paper they receive as a crime scene awaiting investigation. Unless



Robert Laughlin was critical of the selection process for reviewers of Schön's papers.

there are obvious grounds for suspicion, "I never question the authenticity of the data", says Bert Vogelstein, a cancer researcher at Johns Hopkins University in Baltimore, who has a reputation among editors as a thorough referee. "If a clever person wants to manipulate the data, it is hard enough to catch in the lab, let alone in the paper."

But according to researchers who have investigated scientific misconduct, fraudsters aren't always so clever. Ulf Rapp, a cell biologist at the University of Würzburg in Germany, headed the inquiry into one of the most spectacular cases of scientific fraud in recent times: that perpetrated by cancer researchers Friedhelm Herrmann and Marion Brach, who worked at the Max Delbrück Centre for Molecular Medicine in Berlin in the early 1990s. Rapp's task force examined 347 publications and concluded that data in 94 of them had either definitely or "highly probably" been manipulated¹⁴. He now says that, in many of these cases, referees and editors should have spotted the apparent duplication of data. "Some cases were clearly a failure on the part of the reviewer and the journals. They were very superficially evaluated," Rapp says.

Paper copies

Schön's manipulation and fabrication of data was less obvious. But still, had the reviewers of his papers, or the editors who handled them, placed the figures they contained alongside those from work that he had published previously, they might have noticed suspicious signs of data duplication. After all, when the news first broke that several of Schön's papers were under investigation, it didn't take long for physicists to identify more publications containing questionable data.

Although many journals issue guidelines to referees detailing particular sections of the paper that should receive scrutiny, few offer specific advice about how and what to check



Limits of detection: Bert Vogelstein says that clever fraudsters are almost impossible to catch.

Z. KAREEM/JOHNS HOPKINS

and in what level of detail. So, in light of the Schön case, should journals require referees to check the data in papers under review against those from previous publications? And should *Nature's* experience with Quist and Chapela's manuscript encourage journals to demand that referees make more stringent checks on DNA-sequence data to ensure that they support a paper's conclusions?

Most scientists and editors believe that such prescriptive approaches are unlikely to be of much help, given the huge diversity of data in the papers under review. Leo Kouwenhoven, a nanotechnologist at the Technical University of Delft in the Netherlands, believes the Schön case will make researchers in his field take more care in comparing figures with previously published work. But he doubts that this, in itself, will do much to prevent future scandals. "This time it was duplicated figures to look out for, but next time something else will be the problem," he says.

Rather, most experts believe that it would be more fruitful to consider ways to encourage improved diligence in general. Almost all editors interviewed for this article said that, in their experience, the standard of refereeing is extremely variable. "The refereeing process is hit-and-miss, and part of an editor's job is to understand that and choose the right referees," says Ball.

Some referees go to incredible lengths and will even re-plot data to check that a paper's conclusions are correct. In crystallography, for example, some reviewers run structural coordinates through their own computer programs to check the resulting molecular structure against the one described in the manuscript. Others will return a brief paragraph from which it is clear, to an experienced editor, that they have given the paper only a superficial read.

"I think one reason that the quality of refereeing often isn't very good is that it's not

Some referees will even re-plot data to check that a paper's conclusions are correct.

important to the referee," says Simon Wain-Hobson, an AIDS researcher at the Pasteur Institute in Paris. "It's not a priority and people don't break their back over it. I've had colleagues say they can review a paper in ten minutes."

Burden of trust

Skilled editors recognize shoddy refereeing, and tend to ask those scientists who are most conscientious to review more than their fair share of manuscripts. "There is a huge logistical burden, particularly if someone gets a reputation as a thorough reviewer," says Weil.

But can journals provide incentives to encourage improved diligence across the board? One way might be to pay referees, and a few journals have taken this step. The *IBM Journal of Research and Development*, for example, pays a minority of its referees a few hundred dollars per paper. "We do this for reviewers who have submitted a careful, thorough review," says John Ritsko, editor-in-chief of IBM Technical Journals. "It is essentially a 'thank-you' for a job well done."

Paying referees is more common among economics journals. This is partly because — unlike science journals — they regularly publish papers after their results have been widely disseminated, and so need incentives to hurry referees who feel that there is no rush. Referees for the *Journal of Political Economy*, for instance, receive US\$40 if they return their report within three months and \$75 if they finish the task within six weeks. Nevertheless, many referees remain unimpressed by the financial reward and still fail to deliver, says Vicky Longawa, the journal's managing editor — and much of the cost is passed on to authors, who pay \$50 to submit each paper.

Few scientific journal

publishers are enthusiastic about introducing a system of payments that would require the industry's business model to undergo a complete upheaval. And paying referees may not improve the system, says Joshua Gans of the Melbourne Business School in Australia, who is an expert on game theory — which can help to predict people's behaviour when faced with particular reward schemes. The system works at present because academics feel obliged to take responsibility, he argues. Paying referees "can motivate an academic but at the same time take away their feelings of professional responsibility because they know others are motivated by pay, too," Gans claims.

Rewards need not be financial, however. For two decades, the American Geophysical Union (AGU) has run a scheme to honour excellence in refereeing for its journals. Those cited by journal editors get their names and pictures published in the AGU's *EOS* newsletter, as well as a certificate and an invitation to an awards dinner. Judy Holoviak, the AGU's director of publications, says that the system was introduced because some researchers were refusing requests to act as referees or returning one-line reports. "There are still people who will never review a paper," says Holoviak. "But it is probably good for the young people coming in; they see this and hopefully it is setting a standard."

Although such schemes may have some merit, many editors and scientists still feel that the current system should be left essentially unchanged. Vogelstein paraphrases Winston Churchill's quip about democracy: "It's the worst system in the world, except for all the others." Rather than focusing on the problems, suggests Fred Alt, an immunologist and cancer researcher at Harvard Medical School in Boston, we should remind ourselves of how well the system works. "It is impressive that the majority of what is published is reproducible and accurate," he says.

Other scientists argue that it would be impossible to make the system foolproof — and misguided even to try. "You're asking for a judgement, an opinion, and it has to stay that way," says Wain-Hobson. "The most important thing about a paper isn't that it's been peer reviewed but that it's reproducible. The real peer review only starts when it's published." ■

David Adam is a News and Features writer for *Nature* in London; Jonathan Knight is *Nature's* contributing correspondent in San Francisco.

1. Taleyarkhan, R. P. et al. *Science* **295**, 1868–1873 (2002).
2. Quist, D. & Chapela, I. H. *Nature* **414**, 541–543 (2001).
3. *Nature* **416**, 600 (2002).
4. Brumfiel, G. *Nature* **419**, 419–421 (2002).
5. *Nature* **419**, 417 (2002).
6. Kennedy, D. *Science* **297**, 13 (2002).
7. Kennedy, D. *Science* **298**, 495 (2002).
8. Putterman, S. J., Crum, L. A. & Suslick, K. <http://arxiv.org/abs/cond-mat/0204065> (2002).
9. Metz, M. & Fütterer, J. *Nature* **416**, 600–601 (2002).
10. Kaplinsky, N. et al. *Nature* **416**, 601 (2002).
11. Quist, D. & Chapela, I. H. *Nature* **416**, 602 (2002).
12. Bailes, M., Lyne, A. G. & Shemar, S. L. *Nature* **352**, 311–313 (1991).
13. Lyne, A. G. & Bailes, M. *Nature* **355**, 213 (1992).
14. Abbott, A. *Nature* **405**, 871–872 (2000).



International unions concerned about biodata

Action must be taken now to ensure that data are safely archived and always accessible.

Sir — The guaranteed and sustained public availability of primary, fundamental, experimental scientific data is a matter of considerable concern. Such data include (but are not exclusive to) nucleotide sequences of biological organisms, amino-acid sequences of proteins, three-dimensional structures of biological molecules, and other data produced by genomics and proteomics studies.

In Correspondence (*Nature* **417**, 222; 2002), D. Agosti and N. F. Johnson stress the importance of open access to taxonomic data, noting that the situation for basic taxonomic data is much worse than for genomic data. But even for genomic and structural data there are no internationally agreed mechanisms for ensuring continuing open access to data, and no strict rules for their deposition in public archival databases. These pressing issues have recently been considered by the Inter-Union Bioinformatics Group (IUBG), which contains, under the umbrella of the International Council for Science (ICSU), representatives from several international unions: the International Union for Pure and Applied Biophysics, the International Union of Biochemistry and Molecular Biology, the International Union of Pure and Applied Chemistry, the International Union of Crystallography and the ICSU Committee on Data for Science and Technology. The

IUBG report of May 2002 is available at <http://md.chem.rug.nl/~berends/IUBG-FinalReport.html> or via www.IUPAB.org.

In the fields of genomics, proteomics and macromolecular structures, the primary scientific data, which form an essential part of a scientific publication, are not included in detail in publications, but are deposited in databases. It has always been the practice that those who claim scientific advances in their published work support their claim by making the objective data on which their claim is based openly available. Therefore, such data must be available on at least the same basis as the publication itself, if the common standards of scientific integrity are to be maintained.

The databases concerned are at present maintained by institutions that do not have the support status of national libraries. It is not yet generally recognized at government level that the archiving of such data needs protection similar to the archiving of literature; the responsibilities to maintain the collections and safeguard their integrity and access into the distant future are not clearly defined and internationally agreed.

The IUBG report contains four explicit statements and seven recommendations. It recommends: first, that the international scientific unions identify key archival databases and have an active role in standardization; second, that publishers require authors to deposit their primary

data in a key archival database; third, that funding agencies insist on such deposition and actively support primary-data repositories; and fourth, that legislators ensure that laws on intellectual property rights allow the fair use of data for scientific and educational purposes.

The aim of the IUBG report is to stir up the scientific community worldwide. The US government has taken the lead by supporting GenBank and the Protein Data Bank, but the maintenance of archival databases is a supranational activity. At present there are different models for funding various databases and there are different funding models in the United States, Europe and Japan. None has an explicit long-term commitment. The obligation to deposit data must be followed worldwide. There must be a single international archive for each class of data, even if it is distributed over more than one site, and data must remain uniform in format. There is an urgent need for international agreements to stabilize the situation and to guarantee cooperation, consistency and funding.

Jean Garnier*, Herman J. C. Berendsen†

*INRA Centre de Recherche de Versailles, *Unité de Mathématique, Informatique et Genome, RD 10 (route de Saint-Cyr), 78026 Versailles Cedex, France*

†Department of Chemistry, the University of Groningen, Nijenborgh 4, 9747 AG Groningen, The Netherlands

Feeding the world

Sir — Your “Food and the Future” Insight¹ discusses problems of and prospects for agriculture. In the overview article, Anthony Trewavas (pages 668–670 of ref. 1) argues that agricultural technologies have averted and will continue to avert malthusian crises in which the human population exceeds its food supply. Trewavas writes: “The lessons of history are clear. Successive lurches in population number have driven the development of new agricultural technologies designed to provide food for growing populations.”

There are, however, other perspectives. An alternative analysis² shows that the development of new agricultural technologies has been driven by increasing corporatization and economic integration of agricultural processes and products, particularly in the twentieth century when the most spectacular increase in human population size occurred. During this time, famine resulted not from a global or even (according to some

perspectives) local shortage of food^{3–5}, but from poverty and lack of political power among starving people.

Trewavas discusses concerns about how the projected nine billion people that will inhabit the Earth later this century will be fed. Even today's food supply would suffice if cultural preferences could be changed to reduce meat consumption substantially. This change could, in principle, free more than 40% of the world's grain to feed people rather than livestock⁶. But feeding people receives a lower priority in the current food system than does the profit to be made from the global spread of luxury diets — most of which have deleterious effects on both human and ecosystem health⁷.

We require agricultural practices that are more hospitable to native biodiversity than are the industrial methods that prevail today⁸. The three challenges of agriculture are: to feed everyone well; to safeguard biodiversity; and to provide a decent living for those who produce food. These goals are neither incompatible nor

imaginary. From urban gardens in Cuba, to shade-coffee farms in Mexico, to grass-fed beef from Minnesota, to community-supported agriculture supplying food to downtown New Yorkers — some ecologically and economically innovative farmers and consumers are attempting to reshape the food system to emphasize sustainability over production.

Catherine Badgley

Museum of Paleontology, 1109 Geddes Road, University of Michigan, Ann Arbor, Michigan 48109-1079, USA

1. *Nature* **418**, 667–707 (2002).
2. Goodman, D., Sorj, B. & Wilkinson, J. *From Farming to Biotechnology: A Theory of Agro-Industrial Development* (Blackwell, Oxford, 1987).
3. Sen, A. *Resources, Values, and Development* (Blackwell, Oxford, 1984).
4. Drieze, J. & Sen, A. (eds) *The Political Economy of Hunger* (Clarendon, Oxford, 1990).
5. Lappé, F. M., Collins, J. & Rosset, P. *World Hunger: Twelve Myths* (Grove, New York, 1998).
6. Smil, V. *Feeding the World* (MIT Press, Cambridge, Massachusetts, 2000).
7. Gardner, G. & Halweil, B. in *State of the World 2000* (eds Brown, L. R. et al.) 59–78 (Norton, New York, 2000).
8. Daily, G. C. in *Nature and Human Society* (ed. Raven, P.) 104–113 (Nat. Acad. Press, Washington DC, 1997).

The gospel of inevitability

Was the Universe destined to lead to the evolution of humans?

Life Evolving: Molecules, Mind, and Meaning

by Christian de Duve

Oxford University Press: 2002. 336 pp. \$30

Eörs Szathmáry

Towards the end of *Life Evolving*, de Duve states that his message “does not oppose the view that places humankind on top of the tree (for now) and sees the attainment of this position as a significant, perhaps even necessary, step in the unfolding of biological evolution”. From the scientific point of view, his main agenda seems to be to attack what he calls “the gospel of contingency” — in which evolution unfolds as a series of events crucially littered by chance, ultimately leading to extremely improbable states such as the human population — as expressed by George Gaylord Simpson, Ernst Mayr, Stephen Jay Gould and Jacques Monod among others. But there is another agenda: with admirable frankness, de Duve sets out to prove that many hidden or expressed assumptions of religion, illustrated by those of the Catholic Church (the faith in which he was raised), are clearly untenable from the viewpoint of contemporary biology.

According to the Bible, humankind had to come because God saw it fit. According to the adherents of the ‘anthropic principle’, the Universe, for whatever reason, is fine-tuned in such a way that led to evolution, culminating in an intellect that can reflect on it and on itself. The details of evolution are therefore irrelevant. But de Duve does not buy either of these explanations; instead he offers his own, portraying biological evolution as almost determined by itself to climb up towards the human race. The devil is in the details of evolution. In an imaginary play, Lamarck would embrace de Duve, whereas Darwin (I suspect) would not. Setting this aside, does de Duve prove his case? I do not think so.

He offers a wide overview ranging from the origin of life to the ascent of humans, from biotechnology to the autoevolution of the human race. Many of the topics are treated with obvious competence, although I would have appreciated some explanatory diagrams throughout. Yet there are topics where I disagree with de Duve. For example, although it is true that the first RNA could not have made itself (this is true, by the way, for any first unit of evolution), and that we cannot prove for sure that there was an RNA world, the metabolic competence of ribozymes is presented in a way that is biased against the case. Another example: the treatment of the origin of language (announced on the jacket) is too short to capture the

complexity of the problem and misses many of the recent advances in the field. Here de Duve argues cogently that consciousness is a real thing, and offers some kind of an explanation by saying that it “represents some sort of physically energized state and that the particular configuration of neurons in the cerebral cortex serves as generator and supporter of that state”. If this does not strike you as enlightening then you are not alone.

Let me return to the main theme. In a chapter entitled “The Arrow of Evolution”, de Duve distinguishes between two directions of evolution: horizontal and vertical, which roughly coincide with micro- and macroevolution, respectively. The claim is that the first is littered with contingency, whereas the latter is self-guided along certain trends towards increasing complexity. The main reason for this is that in microevolution the population can respond by many alternative mutations to the same selective challenge, whereas in macroevolution the range of genetic changes is severely limited.

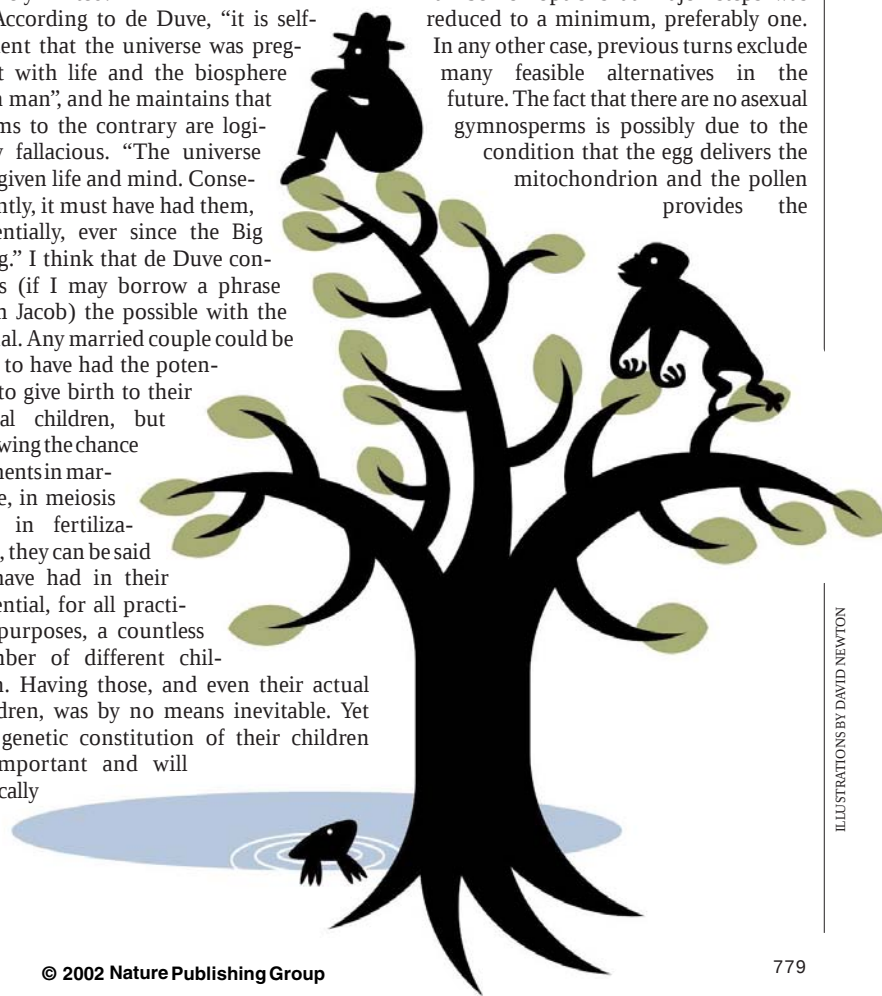
According to de Duve, “it is self-evident that the universe was pregnant with life and the biosphere with man”, and he maintains that claims to the contrary are logically fallacious. “The universe has given life and mind. Consequently, it must have had them, potentially, ever since the Big Bang.” I think that de Duve confuses (if I may borrow a phrase from Jacob) the possible with the actual. Any married couple could be said to have had the potential to give birth to their actual children, but knowing the chance elements in marriage, in meiosis and in fertilization, they can be said to have had in their potential, for all practical purposes, a countless number of different children. Having those, and even their actual children, was by no means inevitable. Yet the genetic constitution of their children is important and will critically

determine many aspects of family life.

I share the belief with de Duve that life had to originate in some more or less straightforward manner, but without a convincing scenario we cannot calculate probabilities and cannot know, therefore, whether this belief is true or not. I am inclined to use contingency exactly opposing de Duve. The point that increasingly complex organisms have relatively few permissible genetic changes (not opposed by negative selection) does not reduce, but instead increases, contingency in the set of possible lineages. Although the alternatives for future evolution are reduced in number, alternatives of existing forms are more markedly delineated from each other, as a result of the constraints of previous turns. To use Lewis Wolpert’s example, one cannot evolve angels with wings from humans, because of the past evolution of the relevant developmental mechanisms.

This increased contingency would lead to inevitability only in the extreme case when, from the level of bacteria upwards, say, the number of options at major steps was reduced to a minimum, preferably one.

In any other case, previous turns exclude many feasible alternatives in the future. The fact that there are no asexual gymnosperms is possibly due to the condition that the egg delivers the mitochondrion and the pollen provides the



ILLUSTRATIONS BY DAVID NEWTON

plastids. The lack of parthenogenesis in mammals may be partly explained by the existence of genomic imprinting. It is not too hard to figure out what genetic changes would be needed to allow for the appearance of asexual variants; it is just highly unlikely that the genetic systems could produce them at once.

Some of the major transitions in evolution (such as the origin of the genetic code, or the eukaryotic cell) could be truly unique, not because of some chance bottleneck, but because either the required genetic variations, or the selective conditions, or both, were extremely unlikely. Given 1,000 Earth-like planets with the same initial conditions, and a period of 8 billion years (to be on the generous side), how many of the planets would evolve eukaryotes? Or deuterostomes, primates or humans? It seems that de Duve's tacit position is that most would have humans. If this case could be proven, it would be the most important discovery in evolution — more important even than the idea of natural selection.

However, de Duve does not offer a proof. Regarding the importance of the asteroid impact at the Cretaceous–Tertiary boundary, he says: “Perhaps mammals were bound to supplant dinosaurs at some stage for reasons linked to the intrinsic properties of the two types of animals, and the asteroid only precipitated an event that would have occurred sooner or later.” Yes, perhaps. But if de Duve wants to supplant divine or anthropic determinism by an evolutionary one, a much stronger case has to be made. Without that, neither physicists, biologists nor the Church will be convinced. ■

Eörs Szathmáry is at Collegium Budapest (Institute for Advanced Study),
2 Szentháromság utca,
H-1014 Budapest, Hungary.

Universal values

The Constants of Nature

by John Barrow

Jonathan Cape/Pantheon Books: 2002.

367 pp. £17.99, \$26

Thanu Padmanabhan

The description of natural phenomena through the differential equations of physics has three separate aspects. First, of course, is getting the correct equations and description. The second has to do with understanding the boundary conditions to these equations without which it is often impossible to select relevant solutions. And finally, one would like to understand the origin of the constants and parameters that occur in the equations.

Although most effort goes into the first task, one should not underestimate the importance of the other two — especially the third. Broadly speaking, one would have expected the number of constants in the equations describing nature to keep reducing as our knowledge increases. This, unfortunately, does not always happen. For example, the fundamental theory describing electroweak interactions has more than a dozen parameters, with their associated constants. One is, therefore, led to questions such as: Why do the constants in the equations have the values they do? Why do dimensionless ratios have some specific values? Are some of the constants of nature more fundamental than others?

These questions, of course, assume that one has a choice in the matter. For example, one might have thought *a priori* that the ratio of the masses of the muon and the electron (which is still undetermined even though we believe we understand the physics of leptons) or the fine-structure constant, a measure of the strength of electromagnetic attractions, could have been ten times as large or as small as observed. Is this really true? One argument — called the anthropic principle — attempts to address this question by stressing that for us to be able to discuss such questions at all, the Universe necessarily evolved in a manner allowing the formation of fairly complex organisms. There have been several attempts to show that if some of the constants of nature had significantly different values, the evolution of the Universe would have been very different and complex organisms could not have originated.

Advocates of the anthropic principle claim that this is the only paradigm currently available to discuss the issue. Opponents criticize the anthropic view for having no predictive power and for introducing a subjective bias (related to the existence of complex organisms) into science.

Another question is whether the constants are truly constant. Laboratory observations cover an insignificant span of time compared with the time over which the Universe has existed in a form familiar to us. If some of the ‘constants’ actually vary with time at a very slow rate, then laboratory experiments cannot determine this, although such a variation can have significant cosmological consequences.

John Barrow discusses these and other related issues in his fascinating book *The Constants of Nature*. In 13 chapters, sprinkled liberally with quotations from many different sources, he discusses the role of constants of nature, the historical quest to understand them, the role of the anthropic principle as a guiding philosophy and some recent evidence suggesting that some of the constants of nature are probably not

constants at all. The major strength of the book lies in the diversity of topics discussed.

Although there are very few equations in the book, it certainly uses a language that is a notch more technical than a non-specialist reader may be accustomed to. For example, graphs are drawn in logarithmic units and the ‘powers of ten’ notation is consistently used to describe large numbers without much explanation.

I found the discussion of the anthropic principle and the description of the theory of the fine-structure constant most impressive. This is to be expected, as Barrow was directly involved in developing these ideas. In a few other places, however, the discussion is somewhat simplistic. For example, in the discussion of the historical evolution of units of measurement, there is no mention of the oriental heritage and discussion is biased towards the ideas of Western civilization. The description of how real advances in understanding physics occur is also far too naive, and the discussion does not merge coherently with the rest of the book. And a typographical error that could be confusing to the reader is a missing factor ‘c’ in the first equation on page 86.

The book is liberally sprinkled with human-interest stories. But to do this properly one should be a historian of science. Otherwise there is a risk of introducing factual errors that have a tendency to propagate themselves. One example here is the mistaken statement that Paul Dirac was the youngest winner of a Nobel Prize for Physics; that honour goes to Lawrence Bragg. Such mistakes rather shake one's faith in the other stories and reduce them to enjoyable bits of gossip, which may or may not be true. ■

Thanu Padmanabhan is at the Inter University Centre for Astronomy and Astrophysics, Pune University, Ganeshkhind, Pune 411007, India.

The last word on Darwin?

Charles Darwin: The Power of Place

by Janet Browne

Alfred A. Knopf: 2002. 624 pp. \$37.50

Ernst Mayr

Apart from the Bible, no book has had a greater impact on the thinking of Western Man than Darwin's *Origin of Species*. This claim is based not merely on Darwin having demonstrated evolution and founded secular science, but also on the fundamental new concepts that he introduced. Darwin's refutation of Aristotle's fourth cause, teleology, was, according to the philosopher Willard Quine, Darwin's greatest contribution to philosophy. Darwin firmly introduced history ("the time factor") into science. In biology, he replaced typological by population thinking and gave chance scientific legitimacy, to mention some of his other contributions. Most major planks in the *Weltanschauung* of a modern thinker can be traced back to Darwin. His role is finally understood, after more than a century of misconceptions and misinterpretations.

In recent years, numerous books with the words Darwin or Darwinism in the title have been published, including several biographies. Even though parts of them were quite satisfactory, they often had serious faults. Their authors tended to promote special ideologies, particularly political ones, resulting in a rather biased product. There was one exception, unanimously praised by all critics: the splendid first volume of Janet Browne's biography of Darwin. As one reviewer, Geoffrey Moorehouse, put it: "If Browne's second volume is as comprehensive and lucid as the first, there will be no need for anyone to write another word on Darwin."

Has Browne fulfilled Moorehouse's hope? She has not overlooked any possible source of information on Darwin, as she spent many years reading Darwin's correspondence (some 14,000 letters) and contemporary journals (her Darwin files contain 347 full reviews of *Origin of Species*, 1,571 general articles and 336 personal large accounts). She also analysed the writings of Darwin's friends and opponents whenever they dealt with Darwin and his work.

Browne offers us a rich account of Darwin's personality, his working habits, his relation to his family, particularly to his children, his interaction with friends and opponents, his non-professional interests (literature and music), his social life, his business activities, and much more. She provides us with a full account of Darwin's puzzling illness, not wasting space on fancy

explanations, such as the South American Chagas' disease, which has completely different symptoms.

I have read no other account of Darwin's personality and daily life that is as rich as Browne's. Four of Darwin's close friends — Charles Lyell, J. D. Hooker, T. H. Huxley and Asa Gray — made major contributions to the ultimate triumph of darwinism. In the United States, Gray was tireless in spreading the gospel of evolution and, although he remained a Christian, he accepted Darwin's paradigm more completely than either Lyell or Huxley, who never accepted natural selection or gradual evolution.

Browne tells the famous story of the arrival of Wallace's manuscript on natural selection and the resulting crisis in Darwin's life. It is well known that Lyell and Hooker published Wallace's letter together with some of Darwin's previously unpublished writings. This deprived Wallace of the sole priority of publication, but presented a factually correct history of the origin of the concept of natural selection. At first, Darwin and Wallace seemed to be in total agreement on the process of evolution, but then Wallace drifted away for various reasons and no longer applied natural selection to the origin of man.

The turmoil surrounding the publication of *Origin of Species* is vividly described by Browne. Quite rightly, it was referred to as "the book that shook the world". The publication of *The Descent of Man* 12 years later

caused much less excitement, even though Darwin was now quite bold in his claims. The response indicated that now the concept of evolution had been largely accepted. No human fossils were known at the time, apart from a few Neanderthal skulls, so Darwin concentrated on demonstrating an animal origin of those human characteristics that are generally considered most typically human, such as language, intelligence, morality, self-consciousness, the religious sense and imagination. Where Darwin had no other argument he fell back on sexual selection. Darwin found that this new principle, which supplemented natural selection, enabled him to explain various phenomena, such as differences between human races, which were hitherto inexplicable.

Natural selection can function properly only if there is abundant genetic variation at all times. Yet in this pre-genetics period there was no understanding of the sources of this variation. This worried Darwin all his life. He continued to gather as much information about it as he could, and continued to propose various theories now known to be invalid. Like many others, Darwin accepted an inheritance of acquired characters as a possible source of new variation. This lamarckian theory was finally and decisively refuted by Weismann in 1883, a year after Darwin's death.

Darwin was extraordinarily versatile in his interests and research. Browne discusses his minor publications, such as *The Expression of Emotions in Man and Animals*



(1872), *The Formation of Vegetable Mould, Through the Action of Worms* (1881) and five botanical publications. J. B. S. Haldane insisted that Darwin's highly original botanical work was as important as his publications on evolution. In some of this research, Darwin showed what an active and successful experimenter he was.

When evaluating this book we must remember that Browne is a historian and biographer, not an evolutionist. This is why she does not feel that it is her job to analyse Darwin's evolutionary paradigm (his five major theories) or to explain the principle of divergence and how it misled Darwin, or why he ultimately failed to solve the problem of the multiplication of species, which had been his major objective when starting to work on his "species book", or to try to explain numerous other evolutionary problems that he encountered but left without explanation. For answers to these questions one will have to turn to other books.

Alas, there still is no satisfactory presentation and analysis of Darwin's whole evolutionary paradigm. My *One Long Argument* (1997) has a short treatment of these problems but, by necessity, does not refer to some of the most recent controversies and findings. To supplement Browne's superb treatment of Darwin, the man and his period, we now need a deep analysis of his work. But this requires a real understanding of evolution, and such an understanding is not very common. ■

Ernst Mayr is at the Museum of Comparative Zoology, Harvard University, Cambridge, Massachusetts 02138, USA.

Reaching for the Moon

The Lunar Men: Five Friends Whose Curiosity Changed the World
by Jenny Uglow
Farrar, Straus & Giroux: 2002. 608 pp.
\$30, £25

Robert Bud

Looking back to the second millennium AD from a great distance in time, the Industrial Revolution might be the only distinctively British feature visible. With a shorter perspective but after a generation of revisionist examination, it seems that something historically unique and world-changing did indeed happen first in Britain during the late eighteenth century. Just when Adam Smith was showing how specialization could lead to steady economic growth, massive technological change began to make possible a quite different kind of growth, transforming an agrarian into a

manufacturing economy. There was also a consciousness that revolutionary changes were in the air, though what and which would be the most important in that era of American independence and French regicide remained unclear. This book takes the reader on a journey in that topsy-turvy time by following the details of a group of individual lives.

Your guide is not the universal historian, but history is a broad church, requiring the collaboration of people with many different skills. Some are remarkable analysts, some imaginative lateral thinkers. Others are expert weavers of tales from the fractured evidence of manuscripts and secondary accounts. Jenny Uglow is a brilliant weaver. She has brought a distinct and wonderful contribution to a subject that has been plentifully studied from other perspectives.

The tale of the Lunar Society of Birmingham is well known to the professional historian of eighteenth-century Britain. A loose club of remarkable pioneers living in the West Midlands (and meeting when the Moon was full, to make homeward journeys safer along unlit roads), it included Josiah Wedgwood, founder of the great pottery; Joseph Priestley, 'discoverer' of oxygen; and James Watt of steam-engine fame. It was among the first of that genre of Industrial Revolution associations, which included Benjamin Franklin's American Philosophical Society and the Manchester Literary and Philosophical Society to which the atom-pioneer John Dalton belonged.

One reason for the plentiful studies of the Lunar Society is that its members were both

prolific inventors of labour-saving devices and laborious writers of letters. Despite losses over 200 years, vast archives of their daily correspondence are still to be found, now carefully catalogued. The letters, not all of which are legible to modern readers, mix business news, gossip and all the elaborate dance of competitive men, conscious of their place in history, communicating with friends who are also rivals in the social world. Even if much of the core narrative already exists, Jenny Uglow is the first to draw so expertly on the texture of the correspondence to weave a picture of the relationships of these men and the world they inhabited.

The book draws not just on the manuscripts but also on two generations of scholarship, including the most current. Along with the sense of human relations there is also a judicious allocation of credit, taking account of recent judgements. The work is not a new analysis in the history of science and technology, but this is not exactly popular history either. When the author refers to 1759 as "that year of victories", she presumes a level of cultural familiarity that may not now be universal. Who today can name Britain's victories against the French that year, never mind understand the significance of the battles of Quiberon Bay or Lagos Bay?

At one level the book can be read almost as a novel of the period, with its pointillist detail and telling social comment. But its ambitions are greater. At various points the reader is encouraged to observe the significance of the Lunar Men and their world to



the reading of the Romantic writers such as Samuel Taylor Coleridge. The conclusion almost wistfully points to the end of an optimistic world and the onset of the romantic period that “spurned the arid insights of reason and the denial of innate instincts”. In a sense, therefore, Uglow, who is honorary professor of English and comparative literature at Warwick University, is in a specific dialogue with readers of the Romantics, urging them to look to the context of the Lunar Men and their concerns.

The specificity of the audience points perhaps to a weakness in the book. It will be enjoyed by many even if they do not pick up every allusion, but scientists may be surprised that the author does not put the Lunar Society into its context in scientific history. The Lunar Men were succeeded not just by the Romantic poets, but also by the British Association for the Advancement of Science, founded in 1831 with such a diverse membership of enthusiasts that William Whewell felt compelled to invent a new name to describe them: “scientists”.

Moreover, although the technological meaning of science is well covered — one might even say exaggerated — its cultural function of aiding understanding of a new, complex, urban, multi-ethnic, multi-class world is neglected. What was the place of the Lunar Men as a group in the Industrial Revolution? Uglow will imply their centrality, but does not decisively address the point. So could the book be better? Certainly many of its readers will go away with unanswered questions and a wish to argue with the author. But that is to attest to achievement, not failure. ■

Robert Bud is at the Science Museum, Exhibition Road, London SW7 2DD, UK.

Einstein brought up to speed

The Curvature of Spacetime: Newton, Einstein, and Gravitation by Harald Fritzsch
Columbia University Press: 2002. 341 pp.
\$29.95, £20

Francis Everitt

When Daniel Defoe's fictional castaway Robinson Crusoe finally collected his thoughts on 30 September 1659, on the lonely island where he would remain for 28 years, 2 months and 19 days before returning to England in 1687, the year Newton's *Principia* was published, his first action was to write down in the notebook he had providentially saved from the wreckage a debit-credit account of his state in two columns, Evil and Good. Physicists on their island, no less remote, despite being linked



to the outer world by mobile phone and Internet, might well set down a similar debit-credit sheet. Knowing so much, they might start with just one column, headed Good, but eventually comes the need to list in an Evil column a few of the many things we don't understand.

Harald Fritzsch's *The Curvature of Spacetime* is a time-travel dialogue set in 1996 between three men: Isaac Newton, aged 45 after completing the *Principia*; Albert Einstein at 51, riding the triumph of relativity; and an imaginary modern expert, Adrian Haller. They meet in various German locations familiar to Einstein, and then at Caltech, where Einstein had spent time in the 1930s. As an expository device, the dialogue form is quite successful. It lets Haller teach Einstein and Newton (and us) the current status, experimental and theoretical, of particle physics, drawing the reader into exchanges of view and conflicting ideas more readily than conventional exposition would allow.

The format is problematic historically, however. We find that Newton apparently knows everything about physics up to about 1890, yet is fogged and slow on the uptake about Einstein's work. Newton had faults — Fritzsch deliberately passes over certain 'disagreeable' aspects of his character — but slowness of uptake was not one of them. And is it pedantry to object to Haller's “reminding” Einstein (in a book containing no mention of Ernest Rutherford) that in 1896 “Henri Becquerel found out that the atomic nucleus of uranium is unstable”? This must have been quite a feat before α -particles, β -particles and γ -rays were known, the electron was discovered, or Rutherford had proposed the nucleus.

Taken literally, Fritzsch's title is misleading. Much of the book is about particle

physics, rather than gravity or spacetime. However, this expansion benefits the content; it reveals how far physics has come since Einstein and shows off Fritzsch's considerable gifts as an expositor.

Both Newton and Einstein puzzled over the meaning of mass. Particle physics has failed to illuminate them: the large unexplained range of masses of elementary particles from neutrinos to quarks has only complicated matters still further. Reading Fritzsch on these issues, including his account through Haller of how the Large Hadron Collider at CERN, the European Laboratory for Particle Physics, will (we hope) reveal the Higgs particle and explain mass, leads to the deeper question of what constitutes explanation in physics.

One of Newton's discoveries, often wrongly credited to Galileo, was that in physics, mass fulfils two functions. According to the law of acceleration, mass is the receptacle of inertia; the inverse-square law of gravity makes it the source of gravitation. The masses in these equations are said to be ‘equivalent’. Equivalence is often said to be incomprehensible in Newton's physics but explained in Einstein's. Einstein's falling elevator made inertial and gravitational accelerations indistinguishable; this indistinguishability became a principle; curved spacetime followed. And so, Fritzsch has Einstein saying that gravitational and inertial accelerations are not only proportional to each other, but “completely identical”.

But surely something is wrong. Complete identity means, in ordinary language, no difference. If there is no difference between inertia and gravity, why does physics need a gravitational constant? Einstein is credited with removing unnecessary concepts — he eliminated the ether, and consolidated

accelerations. His true gift, however, was the unexpected extension of concepts — mass turned into mass-energy, and equivalence became a principle that applies to other phenomena besides mass. His last statement on mass leaves Newton's mystery intact. Distinguishing 'inert mass' from 'heavy mass', Einstein wrote in *Out of my Later Years* (Philosophical Library, 1950): "That these two radically different definitions lead to the same value for the mass of a body is, in itself, an astonishing fact."

Fritzsche's narrative centres on Einstein. Each chapter head has a quotation from Einstein, and Newton rather improbably seems in awe of him. To find Einstein's true greatness requires an act of distancing. Examining the 23 chapter quotations as an entity shows Einstein not to be necessarily all wise, but certainly to be a most unusual person. Consider this: "All I really want is to sit back and find out how God created this world. It is His thought that I am trying to understand — not the spectral lines of this or that element. I really could not care less about things like those." One may disagree. Great discoveries can come from tiny discrepancies — witness the Lamb shift and quantum electrodynamics. But it is hard to think of anyone but Einstein who would have said that. ■

Francis Everitt is at the W. W. Hansen Experimental Physics Laboratory, Stanford University, Stanford, California 94305, USA.

A race through the dark

The Extravagant Universe: Exploding Stars, Dark Energy and the Accelerating Cosmos
by Robert P. Kirshner
Princeton University Press: 2002. 320 pp.
\$29.95, £19.95

Sean Carroll

The 1990s will go down in history as the decade in which we successfully inventoried the constituents of the Universe. A series of ground-breaking observations have given us a picture in which ordinary matter — atoms, stars, gas and dust — accounts for only about 5% of the total energy of the cosmos, with 25% consisting of 'dark matter', which is different from any of the known elementary particles. Most surprising of all, fully 70% of the Universe is the utterly mysterious 'dark energy', characterized by the fact that it is distributed nearly uniformly through space and evolves very slowly (if at all) with time.

The leading candidate for dark energy is Einstein's cosmological constant, which is equivalent to a non-zero minimum amount of energy density at every point in space.

Even though we have measured the amounts of dark matter and dark energy, and can describe some of their simple properties, we know next to nothing about their origin and nature, and are worse than clueless about why they are as abundant as they are. So there is still important work to be done.

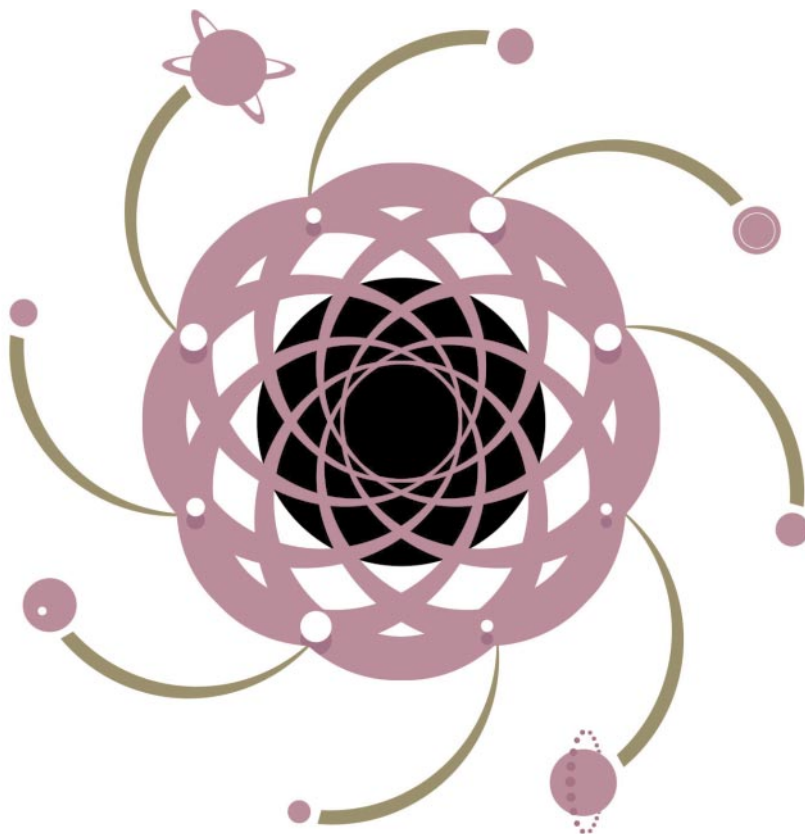
Robert Kirshner played a major role in the discovery that the Universe is accelerating, thereby providing the first direct evidence for the existence of dark energy. In a Universe dominated by ordinary matter, the expansion rate would slow down with time; in contrast, dark energy has a negative pressure, the gravitational effect of which is to make the recession velocity of distant galaxies appear to speed up.

In *The Extravagant Universe*, Kirshner tells the story of how astronomers figured out how to use type Ia supernovae as "standardizable candles" to accurately measure huge distances, and then performed large-scale searches for such objects to determine the behaviour of the Universe's expansion as a function of time. *The Extravagant Universe* is a personal book, rather than an objective account, intermingling the history of cosmology with an explanation of supernovae through the story of Kirshner's own research on these unique celestial events. For the general reader interested in the excitement of how science is done, this strategy makes for a fascinating account. Although several books by leading physicists have recently mixed autobiography with scientific

exposition, most have been by theorists. It is refreshing to get the perspective of someone who has to organize his calendar around the phases of the Moon (in order to leave free those nights when the skies are darkest and best for observation).

Of course, writing a book from a personal viewpoint frees an author from the responsibility of devoting equal amounts of time to the work of all competing groups, although one must still be fair. After a pioneering but premature effort in the 1980s by a group of Danish astronomers, the epochal discovery of universal acceleration in 1998 was the work of two competing teams: the High-Z Supernova Search Team led by Brian Schmidt (of which Kirshner, who was Schmidt's thesis adviser, is a member), and the Supernova Cosmology Project led by Saul Perlmutter. In this mostly friendly rivalry, there is inevitably some jockeying for position in the apportionment of credit; it seems clear, however, that posterity will give both groups full credit for discovering dark energy.

Kirshner is a talented writer, and both experts and general readers will find his book a consistently enjoyable read. He takes the time to tell delightful and surprising stories about the many personalities who have contributed to modern cosmology. You will learn, for example, that the first man-made satellite to escape Earth's orbit was designed by Caltech astronomer Fritz Zwicky, and launched in 1957. (I won't give



away the details.) For my taste, even more enjoyable than the stories is the language itself, which is thick with illuminating metaphors and amusing allusions of all sorts.

For a popular-level book, *The Extravagant Universe* is scientifically ambitious. Kirshner is not afraid to write about stellar spectra in some depth, to use plots of real data, and even to introduce an occasional simple equation. This is heartening, but I think that he might have spent more time trying to explain some of the basic physics at work. There are a number of places where a figure would have made things clearer; for example, although spectral lines are discussed frequently, there is no drawing of a canonical Bohr atom with an electron moving from one level to another. More surprisingly, in a book about the accelerating Universe, there is no plot of the size of the Universe against time, which would demonstrate the difference between accelerating and decelerating.

But these are quibbles, and most readers will learn a great deal from the book, and have fun in the process. The story being told is irresistible in its own right, and is related with verve and good humour. Nobody (or at least, very few people) expected that the Universe would be accelerating, and nobody (literally nobody) knows what dark energy is, or why it has the magnitude it does. Books like this one will help to inspire the next generation of physicists to ultimately answer these questions. ■

Sean Carroll is at the Enrico Fermi Institute, University of Chicago, Chicago, Illinois 60637, USA.

The breath of life and death

Oxygen: The Molecule that Made the World

by Nick Lane

Oxford University Press: 2002. 384 pp. £18.99

Thomas B. L. Kirkwood

The brilliant French chemist Antoine Lavoisier is credited with being the first to demonstrate the importance of oxygen to the world, showing in a famous experiment that the Holy Roman Emperor's diamonds were safe at high temperature provided that oxygen was excluded. Tragically, Lavoisier survived this flirtation with danger only to lose his head to the guillotine in revolutionary Paris on trumped-up charges of false accounting. Over the following centuries, oxygen, named for its 'acid-forming' powers, has become an element widely cherished for its life-supporting properties. We tend to see oxygen in a positive light, with superstar Michael Jackson allegedly spending his

nights in an oxygen tent. Few other elements can claim this level of celebrity endorsement, but the truth is that oxygen is a mixed blessing. Anyone whose life has been marked by fire knows how dangerous it can be, and although we would quickly die without it, growing evidence suggests that oxygen plays a leading part in the ageing process. It is this balance of good and bad that makes oxygen such an interesting molecule.

Nick Lane's enjoyable and informative book would have us believe that "without the threat of oxygen toxicity, life would never have evolved beyond a green slime". Although we depend on it now, it seems highly likely that life first arose in the absence of free oxygen. Oxygen strips organic molecules of electrons and the earliest replicating biopolymers would have been defenceless against its attack. Thus, life began anaerobically and it was during the quiet aeons of early biological evolution that the first photosynthesizing bacteria began to secrete oxygen as a 'toxic' metabolic waste. Not for the last time, the planet's dominant life form made a mess of the environment. This environmental stress imposed a strong selection pressure and there emerged organisms that not only could withstand oxygen toxicity, but which found in this pollutant the source of a new energy supply. Oxidative phosphorylation was entrained in the service of a new breed of life, and, in time, we came along.

In the first half of the book, Lane reviews in some detail the evidence for this role of oxygen in the evolution of life. Although he endorses the major elements of the conventional tale, he prefers a plot line that is relatively new. Instead of a primordial atmosphere made up mainly of methane, ammonia and hydrogen, Lane subscribes to the idea that 4 billion years ago the Earth's atmosphere consisted mostly of nitrogen, as today, with some carbon dioxide and water vapour, and traces of other gases including oxygen. The common view — and mine before reading this book — is that the impressive armoury of defences against oxygen toxicity in present-day organisms evolved to counter the growing danger as free oxygen accumulated in the atmosphere. The alternative, favoured by Lane, is that oxidative stress was known long before free oxygen became a hazard. A significant source of oxidative stress comes from the actions of ultraviolet radiation on water molecules. If life established an early presence in the radiation-exposed ocean surface, where water-splitting photosynthesis might most feasibly have evolved, cells must quickly have acquired potent antioxidant enzymes, such as catalase. So by the time atmospheric oxygen became a threat, the antioxidant defences were already partly in place.

Ultraviolet-induced oxidative stress remains a potent source of free radicals today, attacking any cell exposed to sunlight. Such stress is thought, for example, to play a part in age-related macular degeneration,



one of the most important causes of visual impairment in older people. This brings us to the second half of Lane's book, which is about oxygen's role in killing us. It is a pity that this substantial aspect of the book is not hinted at by the sub-title, because *Oxygen* presents an entertaining and cogent account of how oxidative stress fits into our rapidly expanding knowledge about ageing. Lane also describes work on degenerative conditions such as dementia, and explains with admirable clarity how an imbalance in the antioxidant defence system in people with Down's syndrome, who have an extra copy of the gene for superoxide dismutase, the enzyme which tackles the dangerous superoxide radical but makes harmful hydrogen peroxide in the process, contributes to an increased risk of Alzheimer-like symptoms.

The one shortcoming of the book, perhaps inevitable given its title, is that Lane pushes one molecular player into the limelight, to the exclusion of others, more than is right. The argument that oxygen is the molecule that made the world is hard to swallow, and oxidative stress is not the only agent that makes us age. Nevertheless, Lane presents a nicely crafted account of an important element's place in our lives. His book deserves to be read widely even if, in time, it must share space on the bookshelf with equivalent books on carbon, nitrogen, iron, and the rest. ■

Thomas B. L. Kirkwood is at the Institute for Ageing and Health, University of Newcastle upon Tyne, Newcastle General Hospital, Newcastle upon Tyne NE4 6BE, UK.

Bringing scientists to life

Eurekas and Euphorias: The Oxford Book of Scientific Anecdotes

by Walter Gratzer

Oxford University Press: 2002. 312 pp.

£16.99, \$28

Oliver Sacks

Walter Gratzer has for years been a great stalwart of *Nature*, a man of lucid and encyclopaedic mind whose reviews have moved effortlessly from quantum physics to molecular biology. They are always written with elegance and wit, and are rich with literary and historical associations, for Gratzer seems perfectly at home not just with the concepts and methods of science, but with its history, its personalities, its human aspects. How do scientists come to the thoughts they have, and how may they sometimes delude themselves (and others) on the way? Indeed the subject of self-delusion was the central theme of one of Gratzer's earlier books, *The Undergrowth of Science* (Oxford University Press, 2000).

Eurekas and Euphorias has a much wider scope, and its 181 anecdotes cover the entire range of human genius, folly, character and accident. And anecdotes, properly told, given context and weight, are anything but superficial — some of the anecdotes here are, in effect, miniature essays. (Gratzer quotes Isaac Disraeli, father of Benjamin, who spoke of anecdotes as "minute notices of human nature and of human learning".) Gratzer has assembled a rich collection of historical incidents, conversations, fragments of autobiography and biography, and revelatory fragments, depicting moments of sudden illumination, or paralysing doubt, of intense absent-mindedness and equally intense concentration, accidents which open up wholly unexpected discoveries (there are an astonishing number of these), and incidents that cast a piercing light on the character and work of the scientist. Usually this is a good light, but we also learn some edifying stories of scientific feuds, jealousies, thefts and treacheries, even in individuals as intellectually noble as Newton.

Every anecdote is fully referenced and sourced, and one sees that there is a lifetime of reading behind *Eurekas and Euphorias* — but Gratzer's learning is never intrusive. The range of the anecdotes he has selected defies categorization, although the physical sciences are perhaps more richly represented than the biological ones (thus in the index there are 24 references to Einstein and 18 to Rutherford, but only 9 to Darwin and 7 to Pasteur). Gratzer includes stories of work accomplished under the most adverse

conditions — how Jean-Victor Poncelet's mathematical system was worked out in a Russian prison, and André Bloch's system when he was committed to a lunatic asylum after murdering his family.

There are delicious morsels everywhere. We are introduced to geologist and zoologist William Buckland, the first professor of zoology at Oxford, who would eat anything that came his way — not only mice *en croûte*, or sliced head of porpoise (from animals which had died in the zoo), but on one occasion the embalmed heart of Louis XVI. When shown a stain on a church floor of what was reputed to be the ever-renewing blood of a martyred saint, Buckland bent down, put his tongue to the floor, and diagnosed it instead as bat's urine.

We find Walther Nernst, who conceived the third law of thermodynamics, getting rid of his heat-producing cows because "a thinking man... cultivates animals that are in thermodynamic equilibrium with their surroundings and does not waste his money in heating the universe". We learn of the incredible concentration of Newton, Niels Bohr and David Hilbert; the incredible precocity of Evariste Galois and André-Marie Ampère; the brilliant rudeness of Wolfgang Pauli; the monosyllabic answers of Paul Dirac (if he answered at all). We discover how often accident has played a part, especially in the discoveries of organic chemistry, pharmacology and chemotherapy, as with the chance discovery of the role of mercury in the synthesis of indigo, or the power of mustard gas to combat leukaemia.

Reading these oddly varied, sometimes fortunate, sometimes comic and occasionally dangerous accidents, I could not help

thinking of other similar ones; of how in 1845, for instance, the Swiss-German chemist Christian Schönbein, spilt a mixture of nitric and sulphuric acids in his kitchen lab one day, used his wife's cotton apron to mop it up, and then hung the apron on the stove to dry. There was a flash, a bang — and the apron disappeared. And thus the explosive nitrocellulose gun-cotton was discovered.

But one cannot paraphrase the anecdotes Gratzer tells, for then they become mere one-liners. As Gratzer tells them, these stories have context, dignity and force; they illuminate a man, his work, an era, indeed the whole conceptual and cultural structure of science.

There seems to be no systematic or explicit order to the sequence of these anecdotes, but there is clearly a rightness in their ordering, and this is Gratzer the alchemist, the artist, working invisibly behind the scenes, stringing these anecdotes together like Ezra Pound's *Cantos* or the movements of an opera. Here there is tension, there relaxation; here high seriousness, there levity or farce; here the sublime, there the ridiculous. But equally one can open the book at any point and be educated, thrilled, sobered or surprised, for there is astonishment and delight on every page. I, for one, will put this book next to W. H. Auden's book on aphorisms, John Bartlett's book of quotations, and that ultimate example in illustration, the great Oxford English Dictionary (OED), for finally this is a sort of OED of scientists and science, a banquet of epiphanies, a reference book which is also a work of art.

Oliver Sacks is at 2 Horatio Street 3G, New York, New York 10014, USA and is the author of *The Man Who Mistook His Wife for a Hat*.



A game of chance

Mark Buchanan

When someone makes a 'prediction', he or she usually contends that a specific event will take place at some point in the future: it will rain tomorrow morning; an earthquake will shake the San Francisco area next Thursday. This is 'prediction' as we usually define it. But in modern science, the idea of prediction has evolved subtler interpretations, and the concept continues to develop today, as scientists grapple with phenomena of ever-greater complexity.

By 700 BC, the Babylonians had developed skill in predicting lunar eclipses; they understood their world better and feared it less, presumably, than earlier peoples did. Over the ensuing 2,500 years, science honed its ability to make accurate predictions, with Galileo, Kepler and finally Newton bringing the world's mechanistic predictability to centre stage. In the newtonian Universe, as Pierre-Simon Laplace pointed out, a being of sufficient intelligence could predict the future in detail by knowing the positions and velocities of all particles at any moment.

Today, we casually use the term 'prediction' in a slightly more general sense. One might predict that a new material will be

superconducting below 40 K, or that a mouse that lacks a certain gene will show a particular trait. Such predictions always precede their experimental test, yet aim less to foretell the future than to explore scientific understanding. Prediction in this sense is the engine of science: we design the present (the experiment) and observe the future (the result) so as to compare our theories to empirical reality.

A more interesting twist to the idea arises when forecasting in the newtonian sense is impossible. No one can solve the equations of motion for the 10^{24} particles in a bucket of water, nor would such a solution be useful. For this reason, Boltzmann, Maxwell and other physicists of the nineteenth century executed a strategic retreat to 'prediction' in a weaker sense. One can predict the statistical distribution of molecular velocities in the bucket. Foregoing exact knowledge, statistical mechanics also predicts the likelihood of the 10^{24} molecules being in any particular microscopic state. Strikingly, acquiescence to probabilities at this level yields definite predictions at the higher collective level — we can predict that the water will be solid below 0 °C, for example.

More recently, quantum physics has taken statistical prediction as a fundamental concept, and the idea has also entered the study of chaotic systems, for which detailed prediction is impossible over long periods. Climate scientists refrain from making precise predictions, instead exploring 'scenarios', which effectively project a probable increase in global mean temperature of 1.4–5.8 °C by the year 2100. Earth scientists similarly find it useful to communicate their knowledge to the public by predicting the probability of an earthquake in a given window of space, time and magnitude. Today we are so used to statistical predictions that it is difficult to appreciate the profundity of this retreat into probabilities.

Yet a retreat into statistics does not imply an abandonment of the search for meaningful regularities. Indeed, this search lies at the heart of recent studies of 'complex systems'. Such systems frequently involve many interacting parts that achieved their current organization through a process of evolution. We might think, for example, of a food web, or the intricate tree-like network of streams and tributaries that make up the drainage basin of a great river. In each case, accidental events — such as the chance extinction of a species — have lasting consequences for the system. As with biological evolution, one might replay the 'tape of history' many times and never twice obtain the same result, as the outcome depends on innumerable fine

Prediction

Today we are so used to statistical predictions that it is difficult to appreciate the profundity of this retreat into probabilities.

details that lie beyond scientific scrutiny.

For such systems, one might sensibly wonder whether there is anything to predict. Yet for river drainage basins, studies have identified a well-defined fractal structure — in a statistical sense, river networks exhibit a degree of self-similarity, with small portions of the network being rough copies of larger portions. All drainage basins appear to be alike in this respect, regardless of differences in geography and the many accidents of history that determined their precise structure. Moreover, this 'law-like' feature seems to emerge as the inevitable result of a dynamic process that minimizes the dissipation of energy.

Given this theoretical understanding, it is possible to predict the expected fractal character of any newly discovered basin. An erosion basin recently identified on the surface of Mars appears to validate this prediction. For complex systems, many macroscopic details may indeed be unpredictable; yet predictable statistical regularities may lie hidden behind these details. Recent work on complex networks underlines this point, illuminating deep similarities between the structure of systems as diverse as social networks, the Internet and the biochemistry of the cell.

The idea of prediction may have still further implications for the study of evolution. As the physicist Giorgio Parisi has argued, it may be that for some systems of biological interest — cells, or even large proteins — it will never be possible to predict the properties of individual systems. Science may instead retreat further, making predictions that can be confirmed only by studying the statistics of many cells or proteins, for example. Such an approach may seem to undermine the search for detailed understanding, yet freedom in interpreting the concept of 'prediction' will continue to broaden the scope and power of science. ■

Mark Buchanan is a science writer based in the United Kingdom.

FURTHER READING

Rodríguez-Iturbe, I. & Rinaldo, A. *Fractal River Basins* (Cambridge Univ. Press, 1999).

Albert, R. & Barabási, A.-L. *Rev. Mod. Phys.* **74**, 47 (2002).

Caldarelli, G., De Los Rios, P. & Montuori, M. Published online at <http://xxx.arxiv.cornell.edu/abs/cond-mat/0107228>

Parisi, G. *Physica A* **263**, 557–564 (1999).

LANDSAT/NASA



Twisting tail: the delta of the Lena River in Russia reveals its intricate and convoluted structure.

Atomic photography

Louis F. DiMauro

No ordinary camera can capture the motion of electrons inside an atom. But the advent of ultrafast laser pulses brings the necessary 'shutter speed' for snapping them as they tumble between energy levels close to the nucleus.

In stop-action photography, a moving body is frozen on film by the speed of the shutter or the flash of a strobe light. In this photograph (Fig. 1) of the great American pastime baseball, the shutter speed was sufficient to freeze the motion of the crowd, and the batter's anxious expression, as the pitcher throws the ball towards him. But there is one detail that the camera failed to capture — the motion of the ball. The ball's speed was so high (around 150 kilometres per hour) that it is blurred in the photograph, masking such details as the stitching on its surface. So there are different timescales for the action on the ball field, some too fast to register easily. The same limitation applies at the atomic scale, but the paper by Drescher *et al.*¹, on page 803 of this issue, is set to change this. The authors have introduced a new type of 'photography' that freezes an atom in motion and heralds a new field of research — attophysics.

Consider a water molecule moving through a solution on a picosecond timescale ($1\text{ ps} = 10^{-12}\text{ s}$). On a faster timescale (down to 10^{-14} s), the atomic nuclei of hydrogen and oxygen inside the water molecule also vibrate and bend relative to each other. Scientists have studied this motion since the late nineteenth century by unravelling the mysteries contained in the spectrum of radiation that can be emitted by such a molecule. But the advent of ultrafast laser pulses has revolutionized time-resolved spectroscopy over the past quarter-century.

Using the technique of 'pump-probe' laser spectroscopy, the experimentalist can initiate (pump) and watch (probe) the motion of molecules in real time with a resolution of a few femtoseconds ($1\text{ fs} = 10^{-15}\text{ s}$). This has not only resulted in a direct measure of how molecules move during a chemical reaction but, perhaps more importantly, it has opened up a new school of thought that views the quantum world in terms of moving 'wave packets' representing the evolution of quantum-mechanical amplitudes and phases. In 1999, Ahmed Zewail received the Nobel prize in chemistry for his seminal contributions to this field of femtochemistry.

Although femtosecond light pulses are still an important scientific tool, they cannot freeze every aspect of atomic motion that is relevant in the structure of matter, just as the photograph fails to capture the details of the baseball. The frontier in ultrafast spectroscopy is the study of the motion of

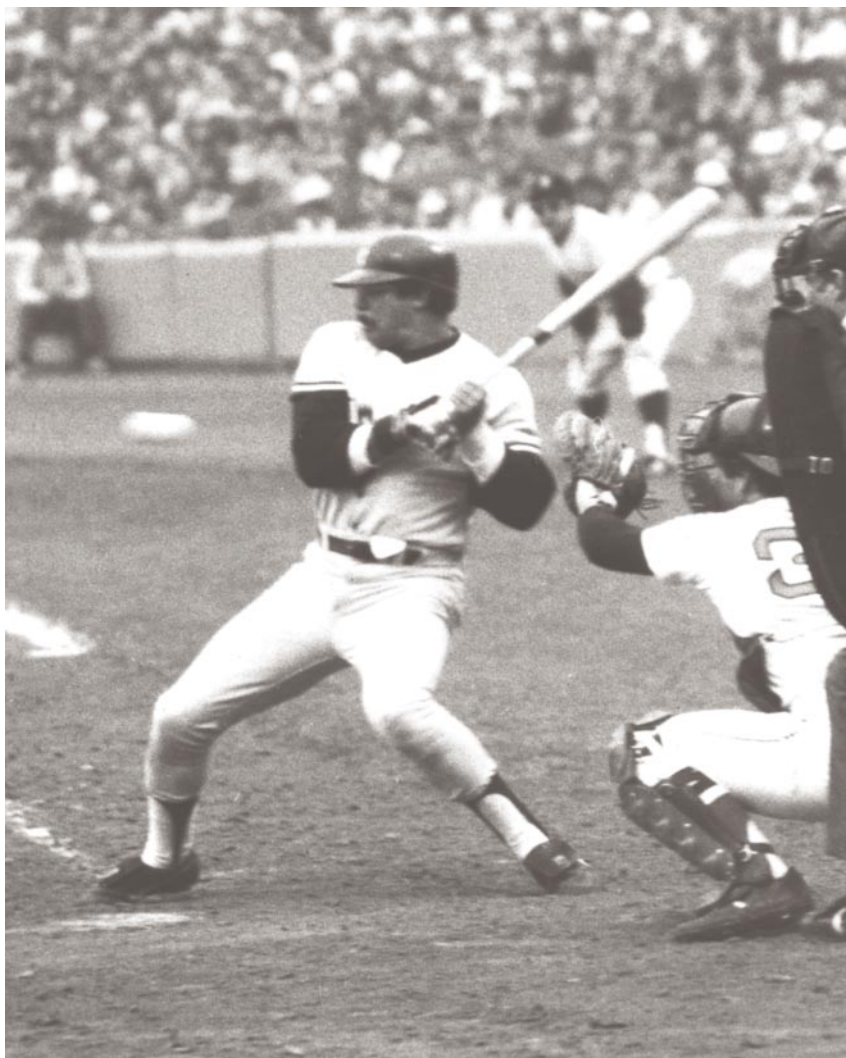


Figure 1 Spot the ball — the limitations of action photography.

electrons bound inside the atom, in orbits close to the nucleus. This new scale is defined by the time it takes the electron in the innermost orbit of the hydrogen atom to complete one turn around the proton nucleus. The period of this orbit — $24 \times 10^{-18}\text{ s}$, or 24 attoseconds — is at least 100 times shorter than the duration of the shortest laser pulse.

But the period of the laser-pulse cycle at or near visible wavelengths imposes a fundamental limit on the resolution that can be achieved, at approximately a few femtoseconds. Consequently, researchers trying to make light pulses of attosecond duration

have looked to other wavelengths, particularly those around the border between the X-ray and ultraviolet (XUV) regimes. One successful approach, using the strong-field phenomenon known as high-harmonic generation, produced the first measurable light pulses in the attosecond regime^{2,3}.

But forming attosecond light pulses is only one of the scientific challenges to be faced in making attophysics a reality. Equally challenging is the detection and propagation of these fragile pulses. Drescher *et al.*¹, however, have solved all of these problems for the specific case of measuring the decay time

of an inner-shell electron excitation, which until now had been studied only indirectly through the emission spectrum.

The authors used an attosecond XUV pulse to initiate the excitation process in an atom of krypton. This pulse frequency is high enough to ionize the atom, knocking out an inner-shell electron and leaving behind an unstable 'hole'. The decay path from this excited state involves two electrons; one drops from a higher shell to fill the hole, and another — the 'Auger electron' — is shaken out of the atom. The Auger decay process occurs between 10^{-14} and 10^{-16} s after ionization, and by sending in a second, longer (femtosecond) optical pulse, Drescher *et al.* could pick up the emission of the Auger electron. The time of release and the final energy of the Auger electron in the intense optical field are related. So by

measuring the Auger electron's energy as a function of the time delay between the XUV and optical pulses, Drescher *et al.* were able to measure the Auger decay time with attosecond resolution.

The emergence of new areas with broad scientific impact is usually the result of creative contributions from a large community of researchers over a long period of time. But there are a few papers that announce the beginning of a new era, and the paper by Drescher *et al.*¹ falls into this category. We are entering a new realm of hyperfast measurement; the age of attophysics has begun. ■

Louis F. DiMauro is at the Brookhaven National Laboratory, Upton, New York 11973, USA.
e-mail: dimauro@bnl.gov

1. Drescher, M. *et al.* *Nature* **419**, 803–807 (2002).
2. Paul, P. *et al.* *Science* **292**, 1689–1692 (2001).
3. Hentschel, M. *et al.* *Nature* **414**, 509–513 (2001).

Cell biology

Survival in three dimensions

Kenneth M. Yamada and Katherine Clark

Whether a cell lives or dies depends on various local cues. New work reveals that those cues include a cell's spatial relationship with its neighbours and polarized interactions with the adjacent extracellular matrix.

Successful cancer treatment requires both the efficient killing of tumour cells and the survival of normal cells. Because cancer chemotherapy involves drug-induced apoptosis (programmed cell death)¹, a crucial question is what actually determines susceptibility to this process or protection from it. As described in *Cancer Cell*, Weaver

*et al.*² have identified a mechanism that is relevant to both normal and cancer cells, in which — surprisingly — the determinants are the three-dimensional organization of the cells and their interactions with the surrounding extracellular matrix.

Most human cancers arise from epithelial cells, which cover body surfaces and the

interior of organs such as the breast, lung and intestine. Epithelial cells are normally anchored to an extracellular matrix called the basement membrane, which gives them polarity. They are organized into three-dimensional structures such as glands. This tissue organization is disrupted when malignant cancer cells invade across basement membranes and into other tissues. Indeed, the ability of tumour cells to thrive in the absence of contact with basement membranes or other such substrates is a well-known characteristic of malignancy³. Weaver *et al.*² now show that these features also make cancer cells vulnerable to drug-induced apoptosis.

In a three-dimensional cell-culture model system, there is a marked difference in tissue organization between a non-malignant breast epithelial cell line and a malignant cell line derived from it. When grown with components of basement membrane, the non-cancerous cells form small, gland-like structures surrounded by locally secreted basement membrane (Fig. 1a, d). In contrast, the tumour-cell counterpart fails to form glandular structures (Fig. 1b). Weaver *et al.* found that treatment with a variety of drugs or immunological inducers of apoptosis killed the tumour cells, but not the well-organized normal cells.

The authors then tested whether three-dimensional organization was crucial to these outcomes. They disrupted the organization of epithelial tissue by growing the cells on flat tissue-culture substrates, producing a single layer (a monolayer). Alternatively, cell-to-cell adhesion could also be disrupted by inhibiting the function of the adhesion molecule known as E-cadherin.

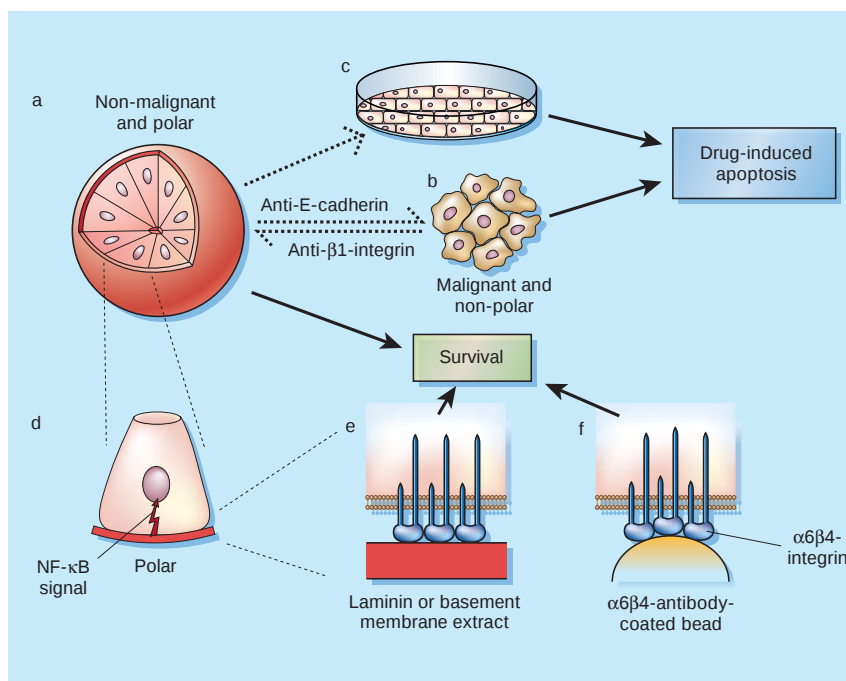


Figure 1 Three-dimensional cell organization and polarity regulate sensitivity to drug-induced apoptosis — a summary of the results of Weaver *et al.*². **a**, Normal, non-malignant breast epithelial cells have polarity and are anchored to a basement membrane (red); they are resistant to drug-induced apoptosis. **b**, Malignant breast cells lack polarity and are susceptible to drug-induced apoptosis. **c**, Converting normal cells to a monolayer by growing them on a flat collagen-coated plastic dish, or disrupting epithelial organization using anti-E-cadherin antibodies, produces cells that are sensitive to apoptosis. Conversely, treatment of breast tumour cells with anti-β1-integrin antibodies induces polarized organization and protects the cells against apoptosis. **d**, Protection is mediated by the NF-κB signalling pathway, which is active in epithelial cells that possess polarity. **e**, **f**, Key molecules involved in this process are laminin in basement membranes and an adhesion receptor, the α6β4-integrin. Cells are also protected from apoptosis by the addition of laminin or a basement-membrane extract (**e**) or by binding to beads coated with an antibody that mimics laminin by binding to the α6β4-integrin (**f**).

Both methods made the cells sensitive to drug-induced apoptosis (Fig. 1b, c). Conversely, manipulation of the malignant cell line through another adhesion system (with $\beta 1$ -integrin antibodies) caused a partial reversion to normal glandular structures, which then were resistant to apoptosis (Fig. 1a, b). Additional experiments showed that cell growth did not affect their sensitivity to cell death.

A key mechanism appeared to be the binding of a specific cell-adhesion receptor, $\alpha 6 \beta 4$ -integrin, which is present on the epithelial-cell surface, to the basement-membrane protein laminin (Fig. 1e). Mutated receptors with abnormal anchorage to the basement membrane did not protect against apoptosis. Interestingly, however, cells were even protected by minimal basement-membrane-type contacts induced by small beads coated with anti- $\alpha 6 \beta 4$ antibodies (Fig. 1f). This finding is important, because it implies that protection occurs at the level of the individual cell: as long as cells can initiate normal $\alpha 6 \beta 4$ -integrin interactions to achieve molecular polarity, they do not need to be organized into glandular structures in order to resist apoptosis. But what links integrin-mediated adhesion to protection from apoptosis? In further experiments, Weaver *et al.*² showed that the key signal was NF- κ B, a well-known regulator of gene expression, thereby linking integrin binding to gene regulation (Fig. 1d).

These findings might help to explain some puzzling observations. For instance, tumour cells growing in monolayer culture are quite susceptible to chemotherapeutic drugs, but become resistant when grown as cell clusters or spheroids suspended in medium⁴. Given the results of Weaver *et al.*, it would be expected that the cells should be resistant to such drugs if they form organized three-dimensional structures. Moreover, a laminin peptide that promotes the formation of three-dimensional structures by salivary epithelial cells can also promote experimental metastasis, perhaps because of its effects on epithelial organization^{5,6}.

A critical question is whether the findings of this new study are directly relevant to human cancers. Is there a direct correlation between the susceptibility of human tumours to chemotherapeutic apoptosis and the absence of $\alpha 6 \beta 4$ -integrin-dependent polarity? If there is, a potentially valuable approach would be to consider $\alpha 6 \beta 4$ -integrin-mediated adhesions as new targets for chemotherapy.

Viewed more broadly, these findings underscore the importance of tissue organization and the three-dimensional relationships of cells. Previous studies established that epithelial-cell polarity is essential for maintaining normal function⁷. Indeed, work with the epithelial-cell model system has shown that three-dimensional organization

affects the regulation of cell growth and tumorigenicity^{8,9}. Conversely, the fibroblastic cells found in connective tissue are normally surrounded by a matrix, rather than polarized towards a basement membrane. Placing fibroblasts in cell culture induces an artificial polarity towards the culture substrate, and returning the cells to a three-dimensional matrix promotes a morphology more like that seen *in vivo*, and stimulates migration and proliferation¹⁰.

These studies with epithelial and fibroblastic cells carry both a warning and a promise. Experiments in which cells are cultured on the commonly used flat surfaces, in which normal three-dimensional relationships with the extracellular matrix and other cells are distorted, may produce mistaken conclusions. Analysing cell interactions in more natural three-dimensional settings, as in Weaver and colleagues' study², promises to provide a view that is closer to what actually happens in living organisms.

Such approaches may also provide a new dimension to cancer treatment, by going beyond the intrinsic genetic, growth and migratory characteristics of cells, and focusing on their microenvironment and organization. ■

Kenneth M. Yamada and Katherine Clark are at the National Institute of Dental and Craniofacial Research, National Institutes of Health, 30 Convent Drive, Bethesda, Maryland 20892-4370, USA. e-mails: kenneth.yamada@nih.gov
katherine.clark@nih.gov

1. Johnstone, R. W., Ruefli, A. A. & Lowe, S. W. *Cell* **108**, 153–164 (2002).
2. Weaver, V. M. *et al.* *Cancer Cell* **2**, 205–216 (2002).
3. Frisch, S. M. & Screaton, R. A. *Curr. Opin. Cell Biol.* **13**, 555–562 (2001).
4. Durand, R. E. & Olive, P. L. *Methods Cell Biol.* **64**, 211–233 (2001).
5. Hoffman, M. P. *et al.* *J. Biol. Chem.* **273**, 28633–28641 (1998).
6. Yoshida, Y. *et al.* *Int. J. Oncol.* **18**, 913–921 (2001).
7. Roskelley, C. D., Srebrow, A. & Bissell, M. J. *Curr. Opin. Cell Biol.* **7**, 736–747 (1995).
8. Weaver, V. M. *et al.* *J. Cell Biol.* **137**, 231–245 (1997).
9. Wang, F. *et al.* *Proc. Natl Acad. Sci. USA* **95**, 14821–14826 (1998).
10. Cukierman, E., Pankov, R., Stevens, D. R. & Yamada, K. M. *Science* **294**, 1708–1712 (2001).

Planetary science

Earth's lunar attic

Clark R. Chapman

Rocks blasted long ago from the surface of Earth and other planets may be preserved on the Moon. Although hard to identify, they could hold a unique record of the chemical history of the planets and even evidence of life.

Two decades ago, it was realized that a few unusual meteorites in our museums are actually pieces of Mars¹. Isotopic ratios derived from gas trapped inside the rocks indicate their provenance. Meteorites that originated on the Moon were soon recognized, and one that might be from Mercury². Calculations of the interplanetary transport³ of material support the possibility that such rocks are being carried to the Earth, and of course there has been speculation that ancient transport of earthly rocks to Mars, lofted by giant cratering impacts, may have 'seeded' life on that planet — or the other way round.

The chances of finding meteorites from Venus on Earth are poor, partly because of the rarity of impacts large enough to eject material out of the planet's very thick atmosphere. But long ago, the inner Solar System was a much more violent place, and the venusian atmosphere might have been thinner, allowing meteorites to escape. Writing in *Icarus*, Armstrong, Wells and Gonzalez⁴ suggest that the Moon may be a repository for precious, ancient meteorites from other planets, including Venus, Mars and the Earth itself.

Armstrong *et al.*⁴ argue that rocks delivered from other planets to the Moon are specially blessed. For the past 3.8 billion years, the lunar surface has been tranquil compared with the pervasive geological and geochemical evolution of planets such as the

Earth, Venus and Mars. As a result, venusian meteorites delivered to the Moon aeons ago (long before the global volcanism on Venus that largely ended half a billion years ago) might remain and could give insight into the history of Venus during epochs for which evidence in the planet itself has been obliterated.

The paper's charming title, "Rummaging through Earth's attic for remains of ancient life", reveals its theme. The authors propose that, very roughly, 20,000 kg of mainly ancient terrestrial rocks may remain in a typical 10 × 10 km² region on the Moon; around 200 kg are expected from Mars, and just 1–30 kg from Venus. But these exotic rocks can't simply be scooped up. Most would be buried hundreds of metres deep in 'megaregolite' — the massive lunar crustal materials that were blasted around the surface when the Moon's huge impact basins, several hundreds of kilometres in diameter, formed.

It is thought that a quarter of the Moon's several dozen impact basins formed during the short 50-million-year interval between 3.9 and 3.85 billion years ago — termed the Late Heavy Bombardment (LHB)⁵. The remaining three-quarters are even older, and a controversial number of basins, since obliterated, formed earlier still. In most LHB scenarios, the Earth and other terrestrial planets were bombarded by asteroids long

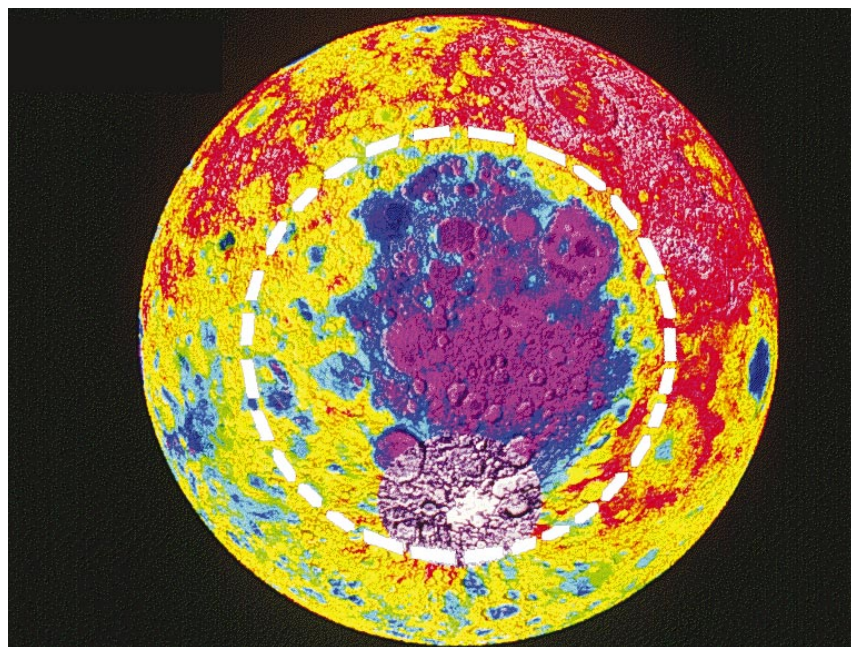


Figure 1 The South Pole–Aitken Basin. This false-colour mosaic of images from the Clementine mission shows the huge impact crater on the far side of the Moon, more than 2,000 kilometres in diameter. The scarring of the Moon records a heavy bombardment by asteroids and comets during the early history of the Solar System. Armstrong *et al.*⁴ suggest that rocks from other planets preserved on the lunar surface would document the chemical history of the Earth and Venus, as well as other planets, during that epoch.

after the Solar System formed (hence the term 'Late'); outer Solar System bodies such as Jupiter's moons were also conceivably bombarded⁶. On Earth, the 100-plus basin-forming impacts would have made for a hellish time indeed, just as life was struggling to gain a foothold on our planet, and possibly on others⁷. Rocks blasted into space from these impacts would dominate the terrestrial 'contamination' of the Moon.

Armstrong *et al.*⁴ estimate that around 7 parts per million of lunar surface materials are of terrestrial origin — in which case there might already be a few grams sprinkled among samples brought back from the Moon. Whether from existing samples or from some future mission to hunt for extra-lunar planetary samples on the Moon, how could such needles in the haystack of lunar soils be recognized? How could every grain be efficiently sampled for a signature of terrestrial or venusian origin?

On the Moon, would these rare rocks even be accessible? We might similarly ask how representative are the lunar samples we already have, and how solid is their evidence for the LHB. Whether returned by astronauts, Soviet robotic missions or even as lunar meteorites, all come from the immediate surface of the well-churned lunar soil (regolith), or from the uppermost few metres. What movement and processing have they been subject to and how certain can we be about which giant impact basin (let alone which planet) they might be from?

Dating the LHB relies on the validity of geological inferences that associate certain rocks with often distant impact basins⁸. Moreover, the spike in impact rates associated with the LHB is actually defined by the rarity of basin-forming impacts before then (inferred from the virtual absence of impact melt-rocks older than 4 billion years⁹). If regolith processes preferentially bury or destroy such older melt-rocks, so that they can't be collected at the surface, then we could be misled. Similar processes might make collecting extra-lunar meteorites even more difficult than it already seems.

Would extra-lunar meteorites even preserve the vital evidence we seek? After all, they must have been blasted off the Earth or Venus in catastrophic explosions, somehow penetrating thick atmospheres, before slamming into the lunar surface at a speed of several kilometres per second. Obviously, such violent delivery might disturb or destroy the valuable biological or chemical information about primordial epochs that we seek.

Understanding the LHB and assessing the practicalities of searching for extra-lunar rocks depend on understanding how impact cratering forms and modifies both the Moon's surface regolith and deeper megaregolith. Armstrong *et al.*⁴ rely on rather simple parametrizations of these processes. But renewed theoretical modelling of regolith processes is under way in several research groups.

The most exciting prospect for under-



100 YEARS AGO

Is there not room for some provisional hypothesis which shall include both Galton's and Mendel's ideas, which are not necessarily antagonistic, but may turn out to be as simultaneously true as the laws of Boyle and Charles, so that the final results may be of the nature of a product or resultant? I mean that instead of drawing a hard-and-fast line between "recessive" and "dominant" characters we may suppose that these differ like heat and cold, in degree but not in kind. ... To take an instance. Last year (1901) I carefully hybridised two varieties of the sweet pea, using lens, paint brush and muslin nets. One variety used was "Gorgeous," of a salmon-orange colour. ... The other variety was a new cream white, Eckford's "Mrs Kenyon," novelty of 1901. ... None of the flowers of the offspring have been cream-coloured; the seeds borne on "Mrs Kenyon" by pollen from "Gorgeous" have all yielded purple flowers unlike either immediate parents, but probably taking their colour from the known remote purple ancestor of our sweet peas. Of seeds borne on "Gorgeous" by pollen from "Mrs Kenyon," eight plants yielded flowers like "Gorgeous" but ten of the plants yielded purple flowers. Here the dominant purple appears to be due to the previous long ancestry; the salmon variety of ten years' standing has several representatives, but not one single cream flower stands for the 1901 novelty. From *Nature* 23 October 1902.

50 YEARS AGO

During the past two years, simple mud-walled, thatch-roofed huts fitted with exit window traps and occupied by volunteer Africans have been sprayed with various formulations of the three insecticides, DDT, BHC and dieldrin, and the effect on the *Anopheles gambiae*, *A. funestus* and culicine mosquitoes entering them observed. Daily mortalities among the mosquitoes were determined by counting the dead individuals on the floors of the huts, and counting those dying later (within 12 hr. of capture) which were caught alive inside the huts in the morning and in the window traps during the night. ... In the later experiments a hut was also treated with the new insecticide, dieldrin. This was applied in the form of a wettable powder at a calculated dosage of 50 mgm. dieldrin per square foot... Results showed this insecticide to be the most efficient of those tested.

From *Nature* 25 October 1952.

standing basin formation and regolith processes is to fly a sample-return mission to the far side of the Moon, near its largest impact crater, the South Pole–Aitken Basin (Fig. 1). Such a mission was given high priority this summer by the US National Research Council's planetary decadal survey¹⁰. Of course, it is doubtful that a first, moderate-cost mission could locate and return extra-lunar meteorites, given their rarity, but it would be a step towards that goal.

Although Armstrong *et al.*⁴ believe that their theoretical calculations are conservative, they could still be too optimistic about the prevalence of extra-lunar materials on the Moon. Such samples must exist, however, in whatever tiny amounts, and set a marvellous challenge as analytical and sampling techniques improve during future decades. ■

Clark R. Chapman is in the Space Studies

Department, Southwest Research Institute, Suite 400, 1050 Walnut Street, Boulder, Colorado 80302, USA.

e-mail: cchapman@boulder.swri.edu

1. Bogard, D. D. & Johnson, P. *Science* **221**, 651–654 (1983).
2. Palme, H. *Science* **296**, 271–273 (2002).
3. Gladman, B. J., Burns, J. A., Duncan, M., Lee, P. & Levison, H. F. *Science* **271**, 1387–1392 (1996).
4. Armstrong, J. C., Wells, L. E. & Gonzalez, G. *Icarus* **160**, 183–196 (2002); Preprint astro-ph/0207316, <http://arXiv.org>
5. Hartmann, W. K., Ryder, G., Dones, L. & Grinspoon, D. in *Origin of the Earth and Moon* (eds Canup, R. M. & Righter, K.) 493–512 (Univ. Arizona Press, Tucson, 2000).
6. Levison, H. F. *et al.* *Icarus* **151**, 286–306 (2001).
7. Maher, K. A. & Stevenson, D. J. *Nature* **331**, 612–614 (1988).
8. Wilhelms, D. E. *The Geologic History of the Moon* US Geol. Surv. Prof. Pap. 1348 (1987).
9. Ryder, G. *J. Geophys. Res.* **107** (2002) (doi: 10.1029/2001JE001583).
10. National Research Council *New Frontiers in the Solar System: An Integrated Exploration Strategy* (2002); www.nationalacademies.org/ssb/newfrontiersfront.html

Ageing

The old worm turns more slowly

Thomas B. L. Kirkwood and Caleb E. Finch

Detailed studies of cellular changes in ageing nematode worms show that they, like humans, suffer progressive muscle deterioration. Randomness of cell damage is another shared hallmark of the ageing process.

When the nematode worm *Caenorhabditis elegans* was first considered as a model for the study of ageing some 20 years ago, few foresaw how valuable it would prove to be. The past decade, in particular, has seen an explosion of work on the genetics of lifespan in this species. Building on pioneering work by Tom Johnson¹ and others, more than 50 mutations that extend lifespan have now been described, and the metabolic pathways regulated by these genes are being unravelled². Some of these genes regulate a key developmental switch, which directs young worms that find themselves in poor environments to adopt a long-lived stress-resistant form (the dauer larva). Other genes control core processes, such as the overall rate of metabolism. These are exactly the kinds of processes predicted to be important for longevity by the evolutionary theories of ageing, in particular the 'disposable soma' theory, which suggests that competition for metabolic resources between processes such as growth, reproduction and cellular maintenance lies at the heart of the ageing process³.

But despite its value in studying the genetics of longevity, a major limitation to the use of *C. elegans* as a model of ageing has been the scarcity of data on the pathology of aged worms. We knew how long worms lived, but not how they died. This situation is radically changed by Herndon *et al.*⁴, writing on page 808 of this issue, who give a detailed description of the cellular changes that occur

with ageing in wild-type and long-lived mutant worms.

The adult worm contains just 959 somatic cells (that is, cells not involved in reproduction) and each cell can be identified in terms of its function. Using a combination of fluorescence and electron-microscopic techniques, Herndon *et al.* have characterized the major cellular changes that accompany ageing in a number of important cell types. Intriguingly, they found that worms are perhaps not so very different from humans, at least in the sense that an important aspect of their decline into senescence is the pro-

gressive deterioration of muscle, known as sarcopenia. By contrast, the worm nervous system appears remarkably well preserved. If old worms wriggle less, it is not, it seems, because they have forgotten how to do so.

A dramatic feature of the changes described in the aged worms is the seemingly random nature of the large variations between individuals; in other words, the changes appear to be highly stochastic. This is all the more remarkable in an organism that, in so many other respects, is under strict genetic authority. Nematode strains have exceptional genetic uniformity (arising from the fact that they are self-fertilizing hermaphrodites), they are cultured in highly uniform environmental conditions and they have a developmental process of almost clockwork precision. We know, of course, that in humans one of the hallmarks of the ageing process is an increase in variability. Indeed, it has been said of humans that we are all born copies but die originals. This variability in human populations can readily be attributed to the uniqueness of our individual combinations of nature and nurture, but such an explanation does not work for worms.

In fact, variability in worm ageing has been staring us in the face, but has only recently been noticed⁵. The conventional way to plot data on longevity in worms, as in other organisms, is as a survival curve, where the percentage still alive is plotted against age. If we plot exactly the same data in another way, showing instead the statistical distribution of age at death, we see at once the extraordinary variation in lifespan between individual worms, despite the uniformity of their nature and nurture (Fig. 1). The range of lifespan for a given strain is wide and not so very different, in terms of standard deviation expressed as a percentage of the mean, from that of outbred, free-living humans.

Where does this variability in worm ageing come from? Not, seemingly, from their nature or nurture. The answer may be that,

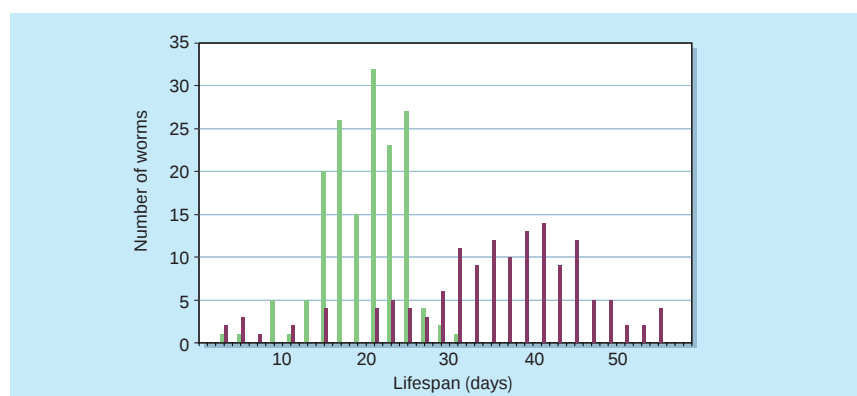


Figure 1 Lifespan distributions for individual *Caenorhabditis elegans* nematodes in isogenic populations of wild-type (green) and long-lived *age-1* (purple) strains. Although the distributions have different mean values, the spread of both (compared to the mean value) is similarly broad. The range of lifespans in nematodes is indicative of the randomness of the ageing process, which Herndon *et al.*⁴ have now investigated at the cellular level (data provided by T. E. Johnson¹).

as predicted by the disposable soma theory, ageing at the cellular level is the result of stochastic damage. Tracking down the molecular changes that underlie the cellular heterogeneity described by Herndon *et al.*⁴ will not be easy, but the first forays into this thicket are being made. It will not be surprising if we find that mitochondrial DNA mutations are important⁶. These are commonly seen as a feature in the ageing of post-mitotic cells such as neurons and muscle across a range of species, including humans.

Equally, we are learning that a range of intrinsic biochemical stresses contribute to the progressive build-up in cells of damage that is, in essence, random. The broad reproducibility of the ageing process from individual to individual masks the fact that at the cellular and molecular level the aged organism is enormously heterogeneous. Neighbouring cells may bear very different burdens of molecular lesions. Genetic differences in ageing rate reflect the fact that the average rates at which damage accumulates are controlled by the efficacy of repair and other maintenance processes.

We know that long-lived *C. elegans* mutants commonly gain their longevity through increasing their resistance to stress⁷. This increase in stress resistance, although it modulates the rate at which stress-induced damage accumulates, will not, by itself, alter the variability embedded in the process. As can be seen in Fig. 1, for the long-lived mutant the range in lifespan measured as a fraction of the mean remains the same as in the wild type. But it remains open whether the random damage accumulated during ageing could account for the large differences between individual worms that are seen even early in adult life. Other stochastic processes may be at work during development of cell contacts, for example in neuromuscular junctions^{8,9}.

Lest we imagine that aged worms can now be taken for a complete model of old people, we must remember that the adult worm is post-mitotic; it has no dividing cells apart from those in its gonads. Thus, *C. elegans* can never be a model for the important contributions to human ageing that come from impaired cell proliferation in the many mammalian tissues and organs that maintain themselves by cell renewal. Nevertheless, the work of Herndon *et al.*⁴ marks a milestone in our advance towards understanding the deep secrets of the ageing process.

Thomas B. L. Kirkwood is at the Institute for Ageing and Health, University of Newcastle upon Tyne, Newcastle General Hospital, Newcastle upon Tyne NE4 6BE, UK.

e-mail: Tom.Kirkwood@newcastle.ac.uk

Caleb E. Finch is at the Andrus Gerontology Center, University of Southern California, 3715 McClintock Avenue, Los Angeles, California 90089, USA.

e-mail: cefinch@usc.edu

2. Hekimi, S., Burgess, J., Bussiere, F., Meng, Y. & Benard, C. *Trends Genet.* **17**, 712–718 (2001).
3. Kirkwood, T. B. L. & Austad, S. N. *Nature* **408**, 233–238 (2000).
4. Herndon, L. A. *et al.* *Nature* **419**, 808–814 (2002).
5. Finch, C. E. & Kirkwood, T. B. L. *Chance, Development and Aging* (Oxford Univ. Press, New York, 2000).
6. Golden, T. R. & Melov, S. *Mech. Ageing Dev.* **122**, 1577–1589 (2001).

7. Lithgow, G. J. & Walker, G. A. *Mech. Ageing Dev.* **123**, 765–771 (2002).
8. Albertson, D. G. & Thomson, J. N. *Phil. Trans. R. Soc. Lond. B* **275**, 299–325 (1976).
9. White, J. G., Southgate, E., Thomson, J. N. & Brenner, S. *Phil. Trans. R. Soc. Lond. B* **275**, 327–348 (1976).

Cancer

Pinning a change on p53

Kevin M. Ryan and Karen H. Vousden

An enzyme-induced conformational change is now implicated in activating the p53 protein, one of a cell's prime movers in preventing tumour development.

One in three of us will succumb to cancer at some point in our lives. So it's not surprising that p53, a protein that helps to prevent the development of malignancies, is the subject of intense scrutiny. In recent years great progress has been made in understanding how p53 is regulated. Two reports in this issue, by Zheng *et al.*¹ and Zacchi *et al.*² (pages 849 and 853), now add a new level of complexity to our knowledge of this regulation by showing that, in order to become fully activated, p53 requires the involvement of an enzyme called Pin1. This discovery adds yet another player to an already crowded field. But it also sheds light on a series of previously described, but poorly understood, phosphorylation events — types of protein modification — that are known to affect p53.

Although it is not needed for normal growth and development, p53 occupies the nodal position in a pathway that monitors and responds to cellular stress, and that has tumour-suppressive effects³. p53 is considered to function primarily as a gene-transcription factor and, once activated, it can induce the expression of a large number of genes, many of which evoke either cell-cycle arrest or programmed cell death (apoptosis).

The type of response depends on the cellular and genetic context, but ultimately the effect is the same — to restrict the proliferation of cells that might otherwise go on to form a malignant tumour⁴.

Because p53 is such a potent inhibitor of cell growth, efficient regulatory mechanisms are necessary to control it during normal cell proliferation. Intrinsic to this regulation is the control of p53 protein stability, mediated by the HDM2 protein (Mdm2 in the mouse)^{5,6}, which keeps p53 levels low in unstressed cells. When p53 function is required, HDM2 activity is blocked, resulting in p53 stabilization and subsequent activation of its tumour-suppressor function.

The correct folding of any protein is fundamental to its ability to function properly. One level at which folding is determined involves the orientation of peptide bonds between adjacent amino acids, which in most cases favours a *trans* configuration — that is, with the two 'central' carbon atoms of the two amino acids lying on opposite sides of the peptide bond, as opposed to the same side (*cis*). Exceptions to this rule are the peptide bonds that link the amino-acid proline with its 'upstream' neighbour (X-Pro), where —

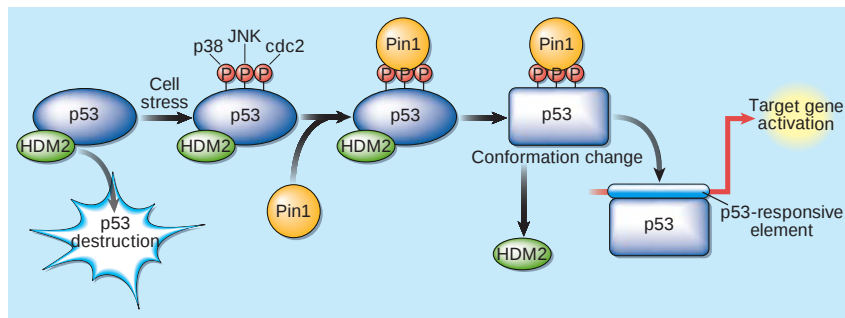


Figure 1 Regulation of p53 protein by Pin1, as suggested by the results of Zheng *et al.*¹ and Zacchi *et al.*². If the cell is not under stress, p53 is not needed and is destroyed by HDM2 (left). Following some cellular insult or other, however, collective phosphorylation of p53 by various kinases — p38, JNK and cdc2 — takes place. The protein is then bound by Pin1, a peptidyl-prolyl isomerase, which changes the conformation of p53, displacing HDM2 and so stabilizing and activating p53. Acting through the p53-responsive element, p53 then functions as a transcription factor for various genes involved in arresting cell growth or inducing programmed cell death. Both processes have the same end, that of preventing tumour formation.

due to lower spatial constraints and a relatively less rigid peptide bond — either the *cis* or the *trans* configuration can be adopted. Switching between *cis* and *trans* forms, which can have significant effects on protein conformation, is catalysed by a group of enzymes named peptidyl-prolyl isomerases⁷.

A special situation exists when the amino-acid X, in X-Pro, is serine or threonine and so is subject to modification by phosphorylation. The addition of a phosphate group can alter the energetic constraints of the Ser-Pro/Thr-Pro bond such that one conformational state is favoured. A recent addition to this scene is the peptidyl-prolyl isomerase named Pin1, which binds specifically to Ser-Pro or Thr-Pro motifs only when the serine or threonine residue is phosphorylated. Pin1 therefore selectively effects conformational change (and possibly also functional change) in proteins that have been targeted by Ser/Thr-Pro-directed kinases, the enzymes responsible for phosphorylation.

Bringing these two fields together, Zheng *et al.*¹ and Zacchi *et al.*² noticed that p53 contains three phosphorylation sites (occurring at serine 33, threonine 81 and serine 315) that are immediately followed by prolines, and so might well be targets for regulation by Pin1. They went on to show that phosphorylation of these sites in response to certain forms of cellular stress results in an interaction between p53 and Pin1 which, through its isomerase activity, brings about a conformational change in p53. The end result was an increase in p53 activity, with both groups finding that loss of Pin1 impairs the p53 response to DNA damage. Intriguingly, both groups show that efficient binding to Pin1 requires phosphorylation at all three sites in p53, implying that Pin1 may integrate signals mediated by different kinases.

Less clear from these studies is exactly how Pin1 affects p53 function. One obvious possibility is that the conformational change induced by Pin1 reduces the interaction of p53 with HDM2, thereby stabilizing the p53 protein and so activating a p53 response (Fig. 1). Indeed, the study from Zacchi and colleagues² provides clear evidence that this occurs in mouse embryo fibroblasts stressed by exposure to ultraviolet light. Nevertheless, Zheng *et al.*¹ found that, in the same cells, although Pin1 was required for p53 activity in response to a different stress signal (doxorubicin), in this case stabilization appeared to be normal. These results imply that, as well as stabilizing p53, the conformational shift induced by Pin1 may contribute directly to p53's ability to function as a transcription factor, possibly by affecting its interaction with DNA or with other proteins. This context dependence is even more dramatically seen in other cells, thymocytes, where the p53 response to γ -ray irradiation was completely unaffected by loss of

Pin1 — suggesting that other peptidyl-prolyl isomerases may be at play in this situation.

But what about the relevance of these observations to tumour suppression? Several mechanisms for regulating the degradation of p53 by HDM2 have now been described, and in some cases perturbation of these pathways has been shown to affect tumour development. So can Pin1 also function as a tumour suppressor? This question is complicated by the ability of Pin1 to target several other proteins, with both positive and negative functions in regulating cell growth and death^{8,9}, but the availability of viable Pin1-deficient mice makes it unlikely that we will wait long for an answer¹⁰. It seems certain, too, that Pin1 will in time add a twist to many other stories, as it has done here for p53. ■

Quantum physics

NOT logic

Nicolas Gisin

A direct quantum equivalent of an electronic NOT gate is impossible. But the best possible approximation to the universal NOT transformation has now been demonstrated using photons.

Quantum mechanics differs radically from classical physics. Accordingly, quantum information processing differs fundamentally from classical information processing. For example, two of the most elementary processes on a classical bit, *b* (equal to 0 or 1), are copying or cloning, and negation — in other words transforming *b* into $1 - b$. But these operations are impossible^{1–3} using quantum bits (qubits) that are a superposition of the basic quantum states $|0\rangle$ and $|1\rangle$. Nonetheless, although they can't be performed precisely, a quantum approximation to these processes should be possible^{3–5}. On page 815 of this issue, De Martini *et al.*⁶ report their investigation of both cloning and negation, and for the latter offer the first near-realization of the universal NOT transformation.

Quantum mechanics has already given us, among other things, the laser and semiconductors, which exploit quantum effects between ensembles of particles. In contrast, the science of quantum information exploits quantum effects at the level of individual particles. There are useful tasks that can be achieved efficiently using quantum information, but that would otherwise be impractical using only classical information processing. This awareness of the potential of individual quantum systems and the related concept of entanglement is at the heart of the growing field of quantum information science.

De Martini *et al.*⁶ used photon polarization for the qubits. For any input polarization, the NOT gate (as in conventional electronics) should transmit the opposite — in this

Kevin M. Ryan and Karen H. Vousden are at the Beatson Institute for Cancer Research, Cancer Research UK Beatson Laboratories, Garscube Estate, Switchback Road, Glasgow G61 1BD, UK. e-mails: k.ryan@beatson.gla.ac.uk k.vousden@beatson.gla.ac.uk

1. Zheng, H. *et al.* *Nature* **419**, 849–853 (2002); advance online publication, 2 October 2002 (doi:10.1038/nature01116).
2. Zacchi, P. *et al.* *Nature* **419**, 853–857 (2002); advance online publication, 2 October 2002 (doi:10.1038/nature01120).
3. Ryan, K. M., Phillips, A. C. & Vousden, K. H. *Curr. Opin. Cell Biol.* **13**, 332–337 (2001).
4. Vousden, K. H. & Lu, X. *Nature Rev. Cancer* **2**, 594–604 (2002).
5. Kubbutat, M. H., Jones, S. N. & Vousden, K. H. *Nature* **387**, 299–303 (1997).
6. Haupt, Y., Maya, R., Kazaz, A. & Oren, M. *Nature* **387**, 296–299 (1997).
7. Shaw, P. E. *EMBO Rep.* **3**, 521–526 (2002).
8. Lu, K. P., Hanes, S. D. & Hunter, T. *Nature* **380**, 544–547 (1996).
9. Rippmann, J. F. *et al.* *Cell Growth Differ.* **11**, 409–416 (2000).
10. Liou, Y. C. *et al.* *Proc. Natl Acad. Sci. USA* **99**, 1335–1340 (2002).

case it should produce a photon in the orthogonal polarization state, for example going from vertical to horizontal polarization, or right circular to left circular. When the photon state is known, quantum NOT gates and cloning are possible, because the result of the transformation can be computed and the corresponding state engineered. But if nothing is known about the input state, this strategy is clearly impossible.

There is also a natural additional constraint, that the quality of the output state should be the same for all possible input states — whatever the polarization of the input photon, the output should be the opposite, the true negation. This property of cloning and the NOT gate is called universality. The polarization states can be represented geometrically as points on a sphere of unit radius, known as the Poincaré sphere. Then, the ideal universal NOT gate should be able to map every point of the sphere onto its diagonally opposite point. It is intuitively obvious that such a transformation is impossible using only rotations. A deeper, related reason why the ideal universal NOT gate is impossible is that it would require an anti-unitary transformation.

How well a quantum process works is defined in terms of 'fidelity' — the ratio of successful processings to the total undertaken. For the optimal universal cloning quantum machine and the optimal universal NOT gate (that is, quantum copying and negation), the fidelities are calculated to be 5/6 and 2/3, respectively^{3–5}. According to quantum theory, these optimal fidelities

should be realizable in optics using stimulated and spontaneous emission — the unavoidable presence of the latter whenever there is stimulated emission explains why perfect quantum cloning is impossible⁷.

De Martini *et al.*⁶ used parametric amplification in a nonlinear crystal to create photons — a technique widely used to produce entangled photon pairs. But, instead of exploiting the spontaneous process of the amplifier, De Martini *et al.* injected a photon into the crystal to stimulate the emission of two similar photons (two clones). When that happened, a third photon was also emitted in another direction. According to theory, this photon, or rather its polarization, represents the best possible approximation of the quantum NOT transformation.

The first experiments to achieve near-optimal quantum cloning have been reported recently^{8–10}. De Martini *et al.* confirm these results, achieving a similar fidelity of 0.810, compared with the theoretical maximum of 0.833, or 5/6. And for the first optimal universal NOT gate, De Martini *et al.* achieve a fidelity of 0.630, close to the theoretical maximum value of 0.666, or 2/3.

Note that the optimal fidelity of 2/3 could also be achieved by simply measuring the input photon in a randomly chosen polarization basis and producing a new photon in a state of polarization orthogonal to the measurement result. As is well known, quantum measurements disturb the system, so even this procedure can't be perfect. But this crude method would achieve optimal fidelity for the NOT gate.

The universal NOT gate demonstrated by De Martini *et al.*⁶, however, is much more in line with today's international effort to develop quantum logic circuits and with the

ambitious goal of building a quantum computer. Whether a futuristic quantum computer will use optics to implement the qubits is unclear. But quantum repeaters for quantum cryptography¹¹ — the means to transmit decoding keys over large distances — will definitely use photons, and the necessary technology will involve techniques similar to the one demonstrated by De Martini and colleagues.

Nicolas Gisin is in the Applied Physics Group, 20 rue de l'Ecole de Médecine, University of Geneva, CH-1211 Geneva 4, Switzerland.

e-mail: nicolas.gisin@physics.unige.ch

1. Wootters, W. K. & Zurek, W. H. *Nature* **299**, 802–803 (1982).
2. Milonni, P. W. & Hardies, M. L. *Phys. Lett. A* **92**, 321–322 (1982).
3. Bužek, V., Hillery, M. & Werner, R. F. *Phys. Rev. A* **60**, R2626–R2629 (1999).
4. Bužek, V. & Hillery, M. *Phys. Rev. A* **54**, 1844–1852 (1996).
5. Bechmann-Pasquucci, H. & Gisin, N. *Phys. Rev. A* **59**, 4238–4248 (1999).
6. De Martini, F., Bužek, V., Sciarrino, F. & Sias, C. *Nature* **419**, 815–818 (2002).
7. Simon, C., Weihs, G. & Zeilinger, A. *Phys. Rev. Lett.* **84**, 2993–2996 (2000).
8. De Martini, F., Mussi, V. & Bovino, F. *Opt. Commun.* **179**, 581–589 (2000).
9. Lamas-Linares, A., Simon, C., Howell, J. C. & Bouwmeester, D. *Science* **296**, 712–714 (2002).
10. Fasel, S. *et al.* *Phys. Rev. Lett.* **89**, 107901 (2002).
11. Gisin, N. *et al.* *Rev. Mod. Phys.* **74**, 145–195 (2002).

Circadian rhythms

Finer clock control

J. D. Alvarez and Amita Sehgal

The clock that governs circadian rhythms is based on a molecular feedback loop, which has just become more complex — two more proteins have been identified as likely components of the loop.

During the 1990s, chronobiologists developed a unified model of the molecular mechanisms that underlie the circadian clock in all organisms. A typical molecular clock consists of an oscillatory feedback loop generated by a few central clock genes. But in order to understand how the clock regulates an organism's physiology and responds to the environment, it is essential to identify other genes and proteins that interact with the core components. On page 841 of this issue¹, Honma *et al.* show that the proteins Dec1 and Dec2 (so-called because they are found in differentiated embryo chondrocytes, or cartilage cells) inhibit transcription

of the main clock genes. Moreover, expression of the *Dec1* gene is induced by light, suggesting that it is involved in mechanisms by which the clock senses the environment.

Circadian rhythms run in the absence of environmental cues, but the central clock nonetheless responds to such cues. For example, light resets the clock, which accounts for the phenomenon of jet lag after an inter-continental flight. At first, one's internal clock keeps time to the original light–dark cycle. It slowly adapts to the new environment and is completely reset after about a week. How light is actually translated into clock alterations is not clear.

Atmospheric science

Plumes and flumes

The fate of smoke billowing from industrial stacks, like those shown here, is a common concern — not least to those who have just hung out their washing. Writing in *Atmospheric Environment* (**36**, 4603–4615; 2002), R. W. Macdonald and colleagues describe modelling investigations of how smokestack configuration with respect to the prevailing wind affects the behaviour of smoke plumes.

As a single hot plume rises, vortices form within it and cool air is drawn in, so reducing plume buoyancy until it rises no further. But when two plumes merge, less cool air is entrained — because of the lower surface-area-to-volume ratio — and the plume is buoyant for longer, rising higher. The distance

between smokestacks clearly influences plume merging. But what about the arrangement of the stacks?

Macdonald *et al.* used a water flume, studying hot-water plumes rising from two 10-cm 'stacks' by repeatedly dragging a grid of temperature probes through the plumes. Plumes in line with the overall flow in the flume quickly mixed and rose higher, as expected: the first plume shields the second from the flow, so that the latter bends less and rises into the first in a way that allows the internal vortices to mingle without destroying overall plume integrity.

Plumes from stacks aligned across the flow did not mix, and rose more slowly — mixing being

hindered because the vortices at the edges of each plume tend to oppose each other. Interaction of vortices on the plume edges can also create a downwash effect, perhaps explaining the slowed plume rise.

Thus, smoke from chimneys set in line with wind direction, rather than across it, is likely to rise higher, travel further, and presumably become more dispersed and diluted before reaching the ground. **Jim Gillon**



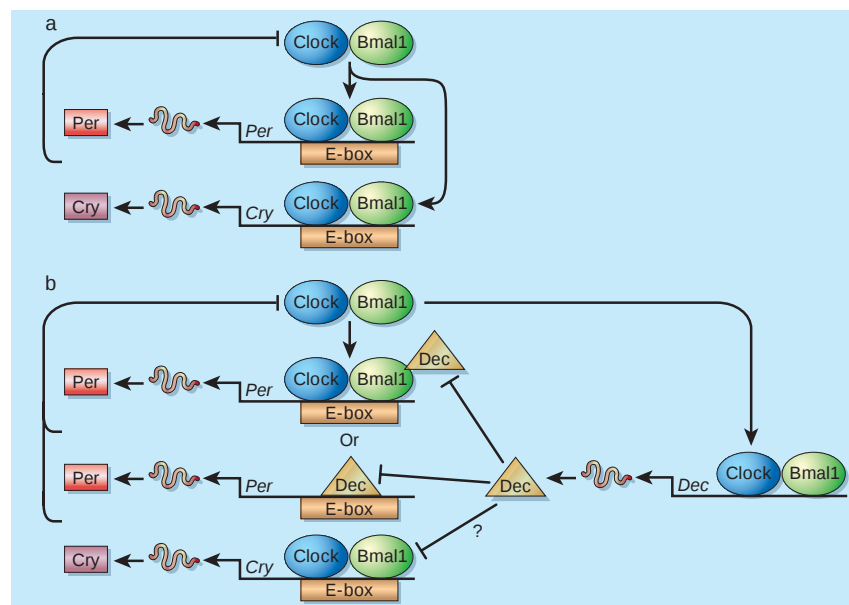


Figure 1 The circadian clock and its regulation. **a**, The molecular basis of the clock is an oscillatory feedback loop consisting of positive and negative components. The positive components are the proteins Clock and Bmal1, which activate transcription of the negative-component genes, *Period* (*Per*) and *Cryptochrome* (*Cry*), by binding DNA sequences called E-boxes. The *Per* and *Cry* proteins repress the activity of Clock and Bmal1. Protein turnover allows the loop to restart. **b**, The Dec proteins, as described by Honma *et al.*¹, repress *Per* transcription by interfering with Clock/Bmal1 activity. The mechanism underlying this interference is not clear — Decs may bind to the E-box or to Bmal1, or to both. It is not known if they repress other genes, such as *Cry*. Transcription of *Dec* genes is reportedly activated by Clock and Bmal1.

In mammals, the central circadian clock resides in a subset of neurons called the suprachiasmatic nucleus (SCN), which lies in the hypothalamus of the brain². It is based on a cyclical feedback loop that includes two proteins, called *Period* (*Per*) and *Cryptochrome* (*Cry*)³. These are negative components because, when their levels are high, both proteins — but *Cry* in particular — repress transcription of the genes for the two proteins, resulting in decreased production of *Per* and *Cry*. The degradation of *Per* and *Cry* over time causes their levels to fall, resulting in relief of the repression and thus restarting the cycle. Both *Per* and *Cry* block the action of two positive clock components, Clock and Bmal1, which activate transcription of the *Per* and *Cry* genes. The overall result is oscillatory expression of clock genes and proteins in the SCN (Fig. 1a). Clock-gene expression also oscillates in tissues other than the SCN, so circadian clocks may occur in most organs of the body. These peripheral oscillators may tailor circadian responses to various physiological conditions, such as hunger and hormone release⁴.

Honma *et al.*¹ propose that Dec1 and Dec2 are components of the core feedback loop (Fig. 1b). These proteins belong to the basic helix–loop–helix (bHLH) family, members of which dimerize with other family members and affect gene transcription by binding to specific DNA sequences called E-boxes⁵. Clock and Bmal1 are also

bHLH proteins and activate transcription by binding E-boxes in the *Per* and *Cry* genes⁶. In contrast, the Decs repress transcription. Structurally, they are related to proteins found in the fruitfly *Drosophila* — *Hairy* and *Enhancer of split* — that function in neural development⁷. Honma *et al.* show that *Dec* expression is cyclic in the SCN and in other brain areas, and that Dec1 and Dec2 can strongly repress Clock/Bmal1 activation of the mouse *Per1* gene (one of three forms of *Period*). Moreover, they report that oscillatory expression of the *Dec* genes occurs in peripheral tissues.

The discovery of cyclic *Dec* transcription in the SCN is compelling. But do these genes function in the core feedback loop? Caution is needed here, because microarray experiments have identified hundreds of genes that show cyclic behaviour in the SCN^{8,9}; presumably, not all of these are components of the central clock mechanism. A finding that favours a central role for the Decs is that only a small subset of genes cycle both in the SCN and in peripheral tissues, as the *Dec* genes do.

The repressive effect of the Decs on circadian transcription suggests that they are negative clock components. The oscillations in *Dec1* and *Dec2* RNA in the SCN lag slightly behind those of *Per1*, which correlates with the repressive effect on *Per1* transcription. If the Decs are indeed repressors of Clock/Bmal1-mediated transcription of *Per1*, then they would seem to be redundant given

that *Per* and *Cry* — most notably *Cry*¹⁰ — have the same function. Perhaps there are extra mechanisms to ensure timely repression of the positive components of the feedback loop. Alternatively, the Decs may regulate the amount of *Per1* synthesized rather than block its expression at a specific time of day. Because *Dec1* expression is upregulated following a light pulse in the middle of the night, it may be involved in the clock's response to light. Notably, *Per1* expression responds similarly to a light pulse. If *Dec1* represses *Per1* expression, it is unclear why both genes would be induced by light. Possibly, *Dec1* limits the *Per1* response and thereby the overall response of the clock to light.

Honma *et al.*¹ have produced some intriguing results. But we're left with many questions. Do the Decs merely repress Clock/Bmal1 activation of *Per1* (and perhaps other Clock/Bmal1 targets), or do they have another function within the clock? Also, how do these proteins accomplish gene repression? The Decs can interact directly with Bmal1, so these proteins may form a large complex on DNA that represses transcription. But the Decs can also bind E-boxes directly, and may inhibit Clock/Bmal1 binding to DNA. Finally, do the Decs function in peripheral clocks, as their oscillatory expression in peripheral organs would suggest? Use of gene-knockout techniques in mice may be needed to answer these questions — work that is probably already under way.

J. D. Alvarez is in the Department of Pathology and Laboratory Medicine, Hospital of the University of Pennsylvania, 6 Founders, 3400 Spruce Street, Philadelphia, Pennsylvania 19104, USA.

e-mail: alvarezj@mail.med.upenn.edu

Amita Sehgal is at the Howard Hughes Medical Institute, Department of Neuroscience, University of Pennsylvania School of Medicine, 232 Stemmler Hall, Philadelphia, Pennsylvania 19104, USA.

e-mail: amita@mail.med.upenn.edu

1. Honma, S. *et al.* *Nature* **419**, 841–844 (2002).
2. Ralph, M. R., Foster, R. G., Davis, F. C. & Menaker, M. *Science* **247**, 975–978 (1990).
3. Reppert, S. M. & Weaver, D. R. *Annu. Rev. Physiol.* **63**, 647–676 (2001).
4. Reppert, S. M. & Weaver, D. R. *Nature* **418**, 935–941 (2002).
5. Massari, M. E. & Murre, C. *Mol. Cell. Biol.* **20**, 429–440 (2000).
6. Gekakis, N. *et al.* *Science* **280**, 1564–1569 (1998).
7. Shen, M. *et al.* *Biochem. Biophys. Res. Commun.* **236**, 294–298 (1997).
8. Ueda, H. R. *et al.* *Nature* **418**, 534–539 (2002).
9. Panda, S. *et al.* *Cell* **109**, 307–320 (2002).
10. Shearman, L. P. *et al.* *Science* **288**, 1013–1019 (2000).

correction

In Russell F. Doolittle's article "The grand assault", commenting on the genome sequence of *Plasmodium falciparum* (*Nature* **419**, 493–494; 2002), the size estimates for the *Dictyostelium discoideum* genome given in Fig. 1 were for its largest chromosome and not the entire genome — which, at around 11,000 genes and 32 megabases, is about four times larger. Labelling on the x axis runs as a log scale 1–10,000.

Obituary

David Keynes Hill (1915–2002)

David Hill, physiologist and biophysicist, died on 18 August 2002. His father was A. V. Hill (known simply as A.V.), a pioneer of biophysics who received a Nobel prize for his measurements of the heat produced by muscles. His mother was Margaret, sister of the economist Maynard Keynes, and of the surgeon and bibliophile Geoffrey Keynes who married a granddaughter of Charles Darwin. A. V. Hill is remembered chiefly for his mathematical formulations, both of the excitation of nerves by applied electric current, and especially of the relations between the force against which a muscle contracts, the speed at which it shortens, and the rates at which it liberates energy as work and as heat.

As a boy, David often acted as an assistant in his father's experiments. Like his father, he was an undergraduate at Trinity, Cambridge, which he entered in 1934 with a major scholarship. In his first two years, he studied physics and chemistry as well as physiology. He then registered as a medical student because in those days a medical qualification was a prerequisite for an academic career in physiology. After doing the necessary anatomy and the final-year physiology course, he had one year of research before the outbreak of the Second World War.

In that year he continued his father's work on isolated muscles from frogs, using for his heat measurements the equipment that A.V. had developed at University College London. He re-examined the absorption of heat that occurs during the first minute after a contraction, and showed that it is caused by the resynthesis of adenosine triphosphate (the immediate source of the energy dissipated during contraction) by transphosphorylation from phosphoryl creatine, and not by a stage in the formation of lactic acid from carbohydrates as had previously been supposed. He also measured the time courses of three other processes during recovery after a contraction that had not been previously measured because of their small size and slow time course: oxygen consumption, production of heat and changes in pH. This was a remarkable achievement for a young scientist in so short a time and on the strength of it David was elected to a research fellowship at Trinity College.

On the outbreak of war, he started clinical study but soon joined a group at the Postgraduate Medical School at the Hammersmith Hospital in London. Here he worked on the effects of crush injuries,



Innovator in muscle research

which were expected to be among the most serious consequences of air raids.

A. V. Hill had led the team developing anti-aircraft gunnery in the First World War, and it was on his advice that in 1940 the physicist Patrick Blackett was appointed as scientific adviser to Anti-Aircraft Command. A.V. then provided him with three physiologists as assistants, including his son David and myself. All of us had an adequate background in physics for our work, which involved adapting the gun-control equipment to operate with the crude data provided by the radar sets of that time, working on a wavelength of 3 metres. This entailed many visits to gunsites, both at practice firings and during air raids.

From 1942, David held various other posts in operational research for the Army, finally becoming personal assistant to Brigadier B. F. J. Schonland, scientific adviser to Field Marshal Montgomery. He was the first British scientist to examine an unexploded V2 rocket, and he obtained information about ballistic missiles that would have been launched by magnetic propulsion into Britain if the launching sites had not been overrun by Allied troops.

After the war, David returned to his research on muscle, first at Cambridge, then (1948–49) as physiologist at the laboratory of the Marine Biological Association in Plymouth, and finally as biophysicist back in Hammersmith (vice-dean, 1970–76). Here his prime duty was overseeing the development of electronics and of instruments for observations on human patients, but he had time to continue his research.

Until 1971 he continued in the same general field as his own and his father's pre-war work, using physical techniques on isolated muscles and nerves but with an emphasis on the intracellular structure of muscle (neglected by A.V.) and using many original techniques. A group of his papers reported optical changes in nerve and muscle. Another series of investigations on muscle used ultraviolet microscopy and radioactive tracers both to localize the adenine nucleotides and phosphoryl creatine in relation to the muscle-fibre striations and to follow the intracellular tubular system that communicates with the external fluid. This system serves to conduct the influence of excitation inwards from the surface membrane of the muscle fibre to activate the contractile material. He discovered a 'short-range elastic component' in the response of resting muscle to stretch, which implied that a number of cross-bridges between the thick and thin filaments exist even in the resting state but are broken when the relative displacement of the filaments exceeds a few nanometres. Almost all of his papers up to 1971 were under his authorship alone.

From 1971 until 1977, David collaborated with R. H. T. Edwards and his group in studies on human muscle, measuring force, heat production and chemical changes, the latter research requiring samples taken from their own muscles by needle biopsy. From 1977, he continued his work on muscle with P. A. Merton, activating their own muscles by heroic procedures such as stimulating them with electric shocks through the skin and by strong magnetic fields applied to the brain.

In 1949, David had married Stella, sister of the immunologist John Humphrey, and they had three daughters. Following his retirement in 1982, David experimented in photography, and, after he and his wife made their final move to Yorkshire, he made himself extremely skilled in woodwork. He died after a long illness in which he became progressively less mobile, though he retained his faculties and remained cheerful until the end.

I got to know David Hill well as an undergraduate: my own career followed his closely from my arrival at Trinity College one year after him. After he left Cambridge in 1948 we kept in touch until his death. His friendship was of great importance to me, especially when I first arrived at Trinity, younger than most of my contemporaries and very shy. He was exceptionally kind, gentle, modest and generous. He will be greatly missed.

A. F. Huxley

A. F. Huxley is at Manor Field, 1 Vicarage Drive, Grantchester, Cambridge CB3 9NG, UK.

Spinning continuous carbon nanotube yarns

Carbon nanotubes weave their way into a range of imaginative macroscopic applications.

The creation of continuous yarns made out of carbon nanotubes would enable macroscopic nanotube devices and structures to be constructed^{1,2}. Here we show that carbon nanotubes can be self-assembled into yarns of up to 30 cm in length simply by being drawn out from superaligned arrays of carbon nanotubes, and that the strength and conductivity of these yarns can be enhanced by heating them at high temperatures. Our findings should help to translate the remarkable mechanical, electrical and thermal properties of carbon nanotubes to a macroscopic scale.

While attempting to pull out a bundle of carbon nanotubes (CNTs) from a CNT array several hundred micrometres high and grown on a silicon substrate, we obtained instead a continuous yarn of pure CNTs (Fig. 1a). This process is very similar to drawing a thread from a silk cocoon, corresponding here to the CNT array. Figure 1b shows a 100- μm -high, free-standing CNT array held by adhesive tape: the indentation at the top of the array marks the region that is being turned into a yarn 30 cm long and 200 μm wide. We estimate that an array area of roughly 1 cm^2 can generate about 10 m of yarn.

Although several methods can be used to prepare CNT arrays on different substrates^{3–6}, we find that continuous yarns can only be drawn out from superaligned arrays in which the CNTs are aligned parallel to one another and are held together by van der Waals interactions to form bundles (Fig. 1c). The yarns usually appear as thin ribbons composed of parallel threads that have diameters in the range of several hundreds of nanometres (Fig. 1d), with the width of the yarn roughly depending on the number of threads in the yarn. In principle, the size of the yarn can be controlled by the tip size of the tool that is used to pick up the yarn — the smaller the tip, the thinner the yarn.

To demonstrate the properties of these yarns, we constructed a light-bulb filament by winding a CNT yarn between two metal leads. The filament emits incandescent light when a DC voltage is applied in a vacuum of 5×10^{-3} Pa (Fig. 1e). After emitting light for 3 h at 70 V, the conductivity of the filament increases by 13% and the tensile strength changes from 1 mN to 6.4 mN. These results indicate that some welding effect may be occurring at the weak connection points of the CNTs during light emission, because these

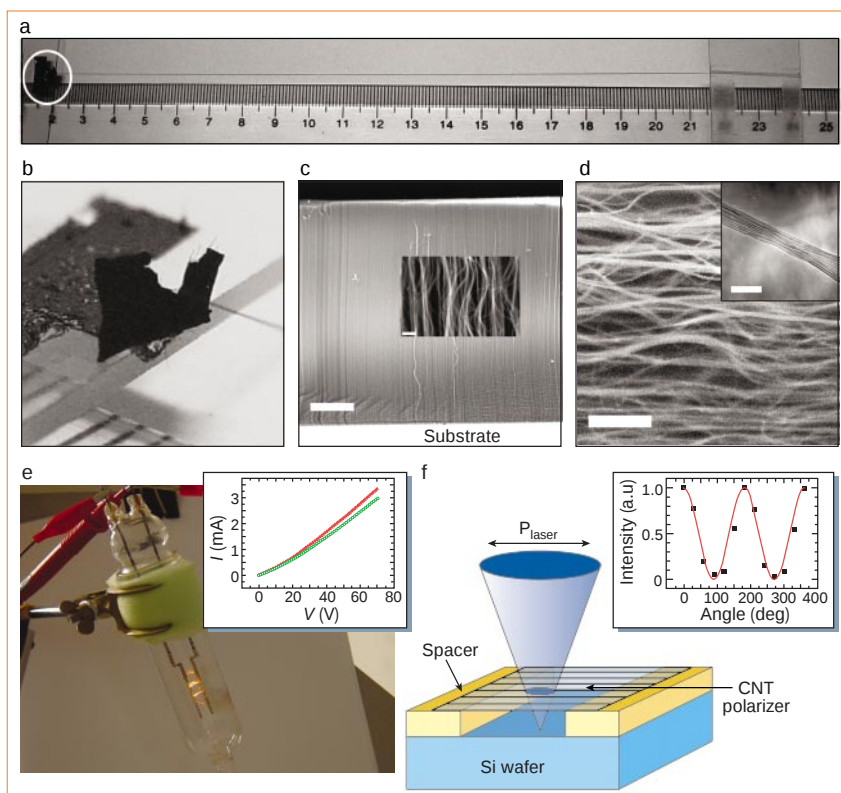


Figure 1 Carbon nanotube yarns. **a**, **b**, A carbon nanotube yarn being continuously pulled out from a free-standing carbon nanotube array (**a**), which is shown enlarged in **b** (roughly $\times 8$ magnification). **c**, Scanning electron microscope (SEM) images of a carbon nanotube array grown on a silicon substrate, showing the superalignment of carbon nanotubes (scale bars: 100 μm ; inset, 200 nm). **d**, SEM image of the yarn in **a**; inset, transmission electron microscope (TEM) image of a single thread of the yarn (scale bars: 500 nm; inset, 100 nm). **e**, Carbon nanotube filaments emitting incandescent light. Inset, I - V curve measured before (green) and after (red) light emission for 3 h at 70 V. **f**, A carbon nanotube polarizer. Polarized laser light (P_{laser}) is focused on the surface of a silicon wafer and reflected back, passing the carbon nanotube polarizer twice before being collected by a CCD detector. Inset, normalized intensity (squares, ratio of the current I to I_{max} ; a.u., arbitrary units) plotted against the angle between the polarization direction of the laser light and the carbon nanotube polarizer, compared with values calculated from Malus's law ($I = I_0 \cos^2 \theta$; red line).

points have a higher resistivity and, as a result, a higher temperature when a current is applied.

We were also able to construct a CNT polarizer by parallel alignment of CNT yarns. When a beam of light passes through the CNT polarizer, photons having a polarization direction parallel to the axis of the CNTs are absorbed, whereas those that are perpendicularly polarized pass through it^{7,8}. Because the CNT diameter is about 10 nm, the polarizer can work in the ultraviolet region, and could even be used for ultraviolet light of wavelengths in the region of tens of nanometres. Figure 1f shows the polarization ability of the CNT device measured at 325 nm, which is in good agreement with that predicted by Malus's law, $I = I_0 \cos^2 \theta$. The degree of polarization of the CNT polarizer, $P = (I_{\text{max}} - I_{\text{min}}) / (I_{\text{max}} + I_{\text{min}})$, is 0.92.

We envisage that pure CNT yarns such as these, particularly after appropriate heat treatment, should eventually be able to be woven into a variety of macroscopic objects for different applications, such as bullet-proof vests and materials that block electromagnetic waves.

Kaili Jiang, Qunqing Li, Shoushan Fan

Department of Physics and Tsinghua-Foxconn Nanotechnology Research Center, Tsinghua University, Beijing 100084, China
e-mail: fss-dmp@mail.tsinghua.edu.cn

1. Vigolo, B. *et al. Science* **290**, 1331–1334 (2000).
2. Ball, P. *Nature* **414**, 142–144 (2001).
3. de Heer, W. A. *et al. Science* **268**, 845–847 (1995).
4. Li, W. Z. *et al. Science* **274**, 1701–1703 (1996).
5. Ren, Z. F. *et al. Science* **282**, 1105–1107 (1998).
6. Fan, S. S. *et al. Science* **283**, 512–514 (1999).
7. Wang, J. F., Gudiksen, M. S., Duan, X. F., Cui, Y. & Lieber, C. M. *Science* **293**, 1455–1457 (2001).
8. Li, Z. M. *et al. Phys. Rev. Lett.* **87**, 1274011–1274014 (2001).

Competing financial interests: declared none.

Olfactory plasticity

One nostril knows what the other learns

About 30% of the adult human population does not perceive an odour when sniffing the steroid androstenone (5- α -androst-16-en-3-one), but will become sensitive to its smell after repeated exposure to the compound^{1–3}. Here we investigate the origin of the plasticity that governs this acquired ability by repeatedly exposing one nostril of non-detecting subjects to androstenone and then testing the unexposed nostril. We find that the exposed nostril and the naive nostril can both learn to recognize the smell, effectively doubling detection accuracy. As the two olfactory epithelia are not connected at the peripheral level, our results indicate that learning occurs in the brain by a mechanism that shares information from both nostrils.

We screened 42 subjects for their ability to detect androstenone by using a four-trial, three-alternative forced-choice paradigm. As expected, this screen yielded 29% non-detectors⁴ (12 out of 42, of whom 7 were female), who were then exposed to this odour for 10 min daily for 21 days. One nostril was blocked by insertion of an inflatable plug, and heated humidified air was injected through the plug at 5 litres min⁻¹ to prevent androstenone from entering the occluded nostril by reverse flow (retronasal olfaction).

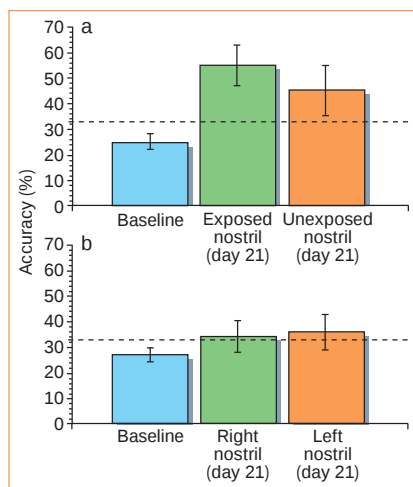


Figure 1 Detection accuracy of androstenone odour by individual nostrils of initially insensitive subjects before (baseline) and after 21 days of exposure. **a**, Experimental group; **b**, control group; dotted lines, chance. Odorants were presented in 60-ml glass bottles to blindfolded subjects. To prevent odour contamination, all testing and exposure was done in a room coated with stainless steel and equipped with high-efficiency particulate air filtration and high-rate carbon filtration. An adjustable vacuum hood was placed over each subject's head. To avoid experimenter-generated cues, all commands (such as "sniff now"), questions (such as "which jar contains the odorant?") and replies were generated by a computer-controlled digitized voice (further details at <http://socrates.berkeley.edu/~borp/supp.htm>).

Before exposure, the experimental group (Fig. 1a) showed 25% accuracy (s.d. 11%), which is not significantly different from 33% chance (binomial from chance: $P = 0.9$). Exposure for 21 days doubled the androstenone-detection accuracy in both the exposed nostril (from 25% to 55%, change from baseline: $t(11) = 3.3$, $P < 0.007$; binomial from chance: $P < 0.002$) and the unexposed nostril (from 25% to 49%, change from baseline: $t(11) = 2.3$, $P < 0.04$; binomial from chance: $P < 0.02$). There was no significant difference in the extent of improvement between the exposed and unexposed nostrils ($t(11) = 0.4$, $P = 0.7$).

We screened a further 50 subjects to obtain 12 more non-detectors (9 female) for a control study (Fig. 1b) in which we tested for confounding possibilities: first, subjects were initially selected for non-detection, and thus their performance could either remain the same or improve, but not deteriorate (a 'floor effect'); second, even a minute leak in the block may have enabled the blocked nostril to learn; third, participation in olfactory testing may improve performance over time. The control study was identical to the experimental, except that both nostrils were blocked.

Control subjects (Fig. 1b) did not differ from the experimental group at baseline (mean accuracy at baseline: experimental group, 25%, s.d. 11; control group, 27%, s.d. 11, $t(22) = 0.44$, $P = 0.7$), and remained at chance detection after 21 days of exposure (change from baseline: $t(11) = 1.6$, $P = 0.14$; binomial from chance: $P = 0.4$). The difference between the two groups after exposure (mean accuracy after 21 days, average for both nostrils) was 52% (s.d. 20) in the experimental group and 35% in the control group (s.d. 14; $t(22) = 2.3$, $P < 0.03$), negating possible confounding factors.

We exploited the paired anatomy of the olfactory system^{5,6} to demonstrate that the plasticity that underpins the emergence of androstenone detection originates in the central components of the olfactory system. These components may be likened to pattern recognition, which occurs at the olfactory bulbs or in primary olfactory cortex — a substrate that shares information from both nostrils⁷ and is optimized for olfactory learning^{8–10}. We do not rule out a contribution to plasticity from the peripheral components of the olfactory system^{11,12} — peripheral receptors may be induced in the unexposed nostril in response to a central signal (direct or hormonal, for example). It remains to be determined how central and peripheral mechanisms could interact to maximize plasticity in the olfactory system.

Joel D. Mainland*, **Elizabeth A. Bremner†**,
Natasha Young*†, **Brad N. Johnson‡**,
Rehan M. Khan*, **Moustafa Bensafi***,
Noam Sobel*††

*Helen Wills Neuroscience Institute, †Department

of Psychology, and ‡Program in Bioengineering, University of California at Berkeley, Berkeley, California 94720, USA
e-mail: mainland@uclink.berkeley.edu

- Wysocki, C. J. *et al. Proc. Natl Acad. Sci. USA* **86**, 7976–7978 (1989).
- Wang, H. W., Wysocki, C. J. & Gold, G. H. *Science* **260**, 998–1000 (1987).
- Dalton, P., Doolittle, N. & Breslin, P. A. *Nature Neurosci.* **5**, 199–200 (2002).
- Gilbert, A. N. & Wysocki, C. J. *Natl Geogr. Mag.* **172**, 514–525 (1987).
- Cain, W. S. *Nature* **268**, 50–52 (1977).
- Sobel, N. *et al. Nature* **392**, 282–286 (1998).
- Wilson, D. A. *J. Neurophysiol.* **78**, 160–169 (1997).
- Haberly, L. B. & Bower, J. M. *Trends Neurosci.* **12**, 258–264 (1989).
- Sullivan, R. M., Wilson, D. A. & Leon, M. *J. Neurosci.* **9**, 3998–4006 (1989).
- Brennan, P. A. & Keverne, E. B. *Progr. Neurobiol.* **51**, 457–481 (1997).
- Nevitt, G. A. *et al. Proc. Natl Acad. Sci. USA* **91**, 4288–4292 (1994).
- Yee, K. K. & Wysocki, C. J. *Physiol. Behav.* **72**, 705–711 (2001).

Competing financial interests: declared none.

COMMUNICATIONS ARISING

Light microscopy

Beyond the diffraction limit

In a comment on resolution in light microscopy¹, Stelzer makes some misleading, if not erroneous, points. His claim to have himself demonstrated the relationship between Abbe's diffraction limit and Heisenberg's uncertainty principle is surprising: this relationship is readily derived from the Fourier theory. The resolution criterion that he seeks to convey is more perplexing: in his description of Dyba and Hell's work², Stelzer ignores the fact that the resolution of any recording system, whether optical or of any other type, is determined by the span of signal frequencies transferred. The higher the transferred frequencies, the finer are the details and the better is the resolution.

A standard way to determine these frequencies and the resolution of a microscope — or of any other signal-recording instrument — is to record the instrument's responses to point- or step-like objects. The responses observed by Dyba and Hell² are clearly of the subdiffraction type, with frequencies beyond the diffraction barrier; they are object-independent by definition. In contrast to Stelzer's view, Dyba and Hell's images, including that of *Bacillus megaterium*, do not imply *a priori* information about the sample. The distance between the bacterial membranes is therefore not relevant for proving that the diffraction barrier has been broken.

Pekka Hänninen

Department of Medical Physics, University of Turku, Turku, Finland
e-mail: pekka.hanninen@utu.fi

- Stelzer, E. H. K. *Nature* **417**, 806–807 (2002).
- Dyba, M. & Hell, S. W. *Phys. Rev. Lett.* **88**, 163901 (2002).

Time-resolved atomic inner-shell spectroscopy

M. Drescher^{*†}, M. Hentschel^{*}, R. Kienberger^{*}, M. Uiberacker^{*}, V. Yakovlev^{*}, A. Scrinzi^{*}, Th. Westerwalbesloh[†], U. Kleineberg[†], U. Heinzmann[†] & F. Krausz^{*}

^{*} Institut für Photonik, Technische Universität Wien, Gusshausstrasse 27, A-1040 Wien, Austria

[†] Fakultät für Physik, Universität Bielefeld, Universitätsstrasse 25, D-33615 Bielefeld, Germany

The characteristic time constants of the relaxation dynamics of core-excited atoms have hitherto been inferred from the linewidths of electronic transitions measured by continuous-wave extreme ultraviolet or X-ray spectroscopy. Here we demonstrate that a laser-based sampling system, consisting of a few-femtosecond visible light pulse and a synchronized sub-femtosecond soft X-ray pulse, allows us to trace these dynamics directly in the time domain with attosecond resolution. We have measured a lifetime of $7.9^{+1.0}_{-0.9}$ fs of M-shell vacancies of krypton in such a pump-probe experiment.

Many-body systems with excess internal energy relax towards states of lower energy by rearrangement of molecular, atomic or nuclear structure. Observing these processes in real time requires a pump pulse for initiating the microscopic dynamics and a delayed probe pulse for detecting transition states of the evolving system. Time-resolved spectroscopy based upon this pump-probe approach¹ is now routinely used for tracking atomic motion in molecules with femtosecond laser pulses².

Electronic motion in weakly bound (Rydberg) states can also be captured with optical or infrared pulses³. However, the relaxation dynamics of core-excited atoms is out of the reach of femtosecond optical techniques. Excitation of a strongly bound electron from an atomic inner shell is followed by an ultrafast rearrangement of the electronic system, resulting in a disappearance of the inner-shell vacancy (henceforth briefly referred to as core hole). Characteristic time constants of the concomitant electronic dynamics range from a few attoseconds ($1 \text{ as} = 10^{-18} \text{ s}$) to a few femtoseconds depending on the energy of the core hole and the strength of electronic coupling.

The lifetime of core holes can be inferred from the energy spectral width of the photons or electrons emitted upon the decay of the excited atomic state. However, except for specific cases of isolated resonances without incoherent background, energy-domain measurements on their own are—in general—unable to provide detailed and accurate insight into the evolution of multi-electron dynamics. As a complementary technique, time-domain access would be desirable. Time-resolved spectroscopy of core-excited atoms calls for several new tools and techniques. The core hole must be created within a time interval that is short compared to its decay time: that is, within typically less than 1 femtosecond. Sampling of the subsequent atomic relaxation calls for a synchronized probe pulse and for a suitable probing technique.

The recent demonstration of isolated sub-femtosecond soft-X-ray pulses^{4,5} and trains of pulses^{6,7} along with attosecond sampling of photo-electron wave packets with a few-cycle optical wave^{8,9} opened the door to attosecond atomic spectroscopy⁹. Here we use these tools and techniques to track rearrangement of the electronic system of a core-excited atom. The core hole is created within less than a femtosecond and its subsequent decay can be observed on a few-femtosecond timescale. The applied technique offers attosecond resolution and will allow time-resolved spectroscopy of a wide range of atomic processes.

Decay of atomic core holes

Excitation of an electron from an atomic shell other than the outermost valence orbital by the impact of an energetic particle—

for example, a photon or electron (here we use photons)—creates a transient hole state of energy W_h (see process a in Fig. 1a). This core hole is not stable, so the system tends to minimize its energy by filling the vacancy with an electron from an outer shell (see process b in Fig. 1a). The excess binding energy $W_h - W_1$ is either carried away by an extreme ultraviolet or X-ray fluorescence photon or transferred via electrostatic forces to another (Auger) electron, which subsequently escapes from the atomic binding (see process c in Fig. 1a). The latter relaxation channel is most probable for light atoms and moderate binding energies ($W_h < 1 \text{ keV}$) but remains intact for $W_h > 1 \text{ keV}$ (ref. 10).

Owing to several Auger relaxation channels, the energy spectrum of the ejected electrons usually contains a number of Auger lines. Further, several photo lines originating from atomic shells with binding energy smaller than the energy of the exciting X-ray photon, $\hbar\omega_X$, add to the spectrum. In Fig. 1a just a single Auger line and a couple of photo lines represent this manifold. The mean kinetic energy of the Auger electron $W_{\text{kin}} = W_h - W_1 - W_2$ and its spectral width Γ are fully determined by intrinsic atomic properties. By contrast, the photo electrons carry the energy of the absorbed photon diminished by their atomic binding energy and hence—in the absence of resonances—their energy distribution reflects that of the exciting radiation.

The feasibility of time-resolved inner-shell spectroscopy relies on X-ray pulses with a duration τ_X much shorter than the lifetime of the generated core hole τ_h . Under these circumstances the temporal evolution of the production rate dN/dt or—in the language of quantum mechanics—the temporal shape $|\psi(t)|^2$ of the ejected wave packet is very different for photo and Auger electrons (Fig. 1b). If the photo-ionization cross-section is nearly constant within the bandwidth of the X-ray pulse, $|\psi_{\text{photo}}(t)|^2$ follows the temporal profile of the exciting X-ray pulse $|E_X(t)|^2$ and thus the production rate of core holes. For $\tau_X \ll \tau_h$, the front edge of $|\psi_{\text{Auger}}(t)|^2$ rises within τ_X , whereas its trailing edge tracks the decay of the core holes. Thus, sampling the time evolution of photo and Auger electron emission explores the generation and decay of core-excited atoms in real time.

Sampling electron emission with light

Strong-field physics provides the key to trace electron emission from atoms. The presence of a strong light field affects the ejected electrons' motion and can be used to probe the emission. Assuming a linearly polarized few-cycle probe light field $E_L(t) = E_a(t) \cos(\omega_L t + \varphi)$, the final momentum (reached after the light pulse left the interaction volume) of an electron released at the

instant t_r is changed by $\Delta p(t_r) = [eE_a(t_r)/\omega_L] \sin(\omega_L t_r + \varphi)$ along the light electric field vector according to a simple classical analysis⁴ (Fig. 2a). E_a is the amplitude envelope, and ω_L and φ stand for the carrier frequency and carrier-envelope phase of the laser pulse, respectively. The final energy is given by:

$$W_f(t_r) = W_i + 2U_p(t_r) \sin^2(\omega_L t_r + \varphi) \cos 2\theta + [8W_i U_p(t_r)]^{1/2} \sin(\omega_L t_r + \varphi) \cos \theta \quad (1)$$

where $U_p(t) = e^2 E_a^2(t)/4m_e \omega_L^2$ is the electron's cycle-averaged quiver energy in the probe light field, e , m_e and W_i are the electron's charge, mass and initial energy, and θ is the angle between its final momentum vector and the light electric field vector. For $\theta \approx 0^\circ$ (Fig. 2b) and $W_i \gg \hbar\omega_L$, a moderate field strength ($U_p \approx \hbar\omega_L$) is able to shift the electron energy by many electronvolts. The energy shift varies from zero to its maximum value within a quarter-wave cycle $T_0/4 = \pi/2\omega_L$ (≈ 0.6 fs at $\lambda_L = 750$ nm). This time-to-energy mapping permits sampling of electron emission with attosecond

resolution. The final energy spectrum for an initial energy distribution $dN/dW = f_w(W_i)$ and time structure $dN/dt = f_t(t_r - t_{r0})$, where t_{r0} represents the instant of the emission peak, can be calculated for any detection direction θ and timing t_{r0} by using equation (1). Experimentally, t_{r0} can be varied by varying the arrival time, that is, the delay Δt (Fig. 2b), of the probe light pulse relative to the X-ray pump pulse that triggers the electron emission. For large initial kinetic energy, $W_i \gg \hbar\omega_L$, the validity of this simple classical treatment of the influence of a strong light field on electrons emitted within a fraction of the light wave cycle T_0 has been corroborated by several photo-emission experiments^{5,8} as well as by quantum mechanical analyses^{11,12}.

We have generalized one (ref. 12) of these quantum theories addressing photo-electron emission in the presence of a strong laser field to describe the behaviour of Auger electrons emitted under similar conditions (for more details see Supplementary Information). Our theory yields the temporal evolution of hypothetical single-line Auger electron spectra probed by strong few-cycle light for different decay times τ_h as shown in Fig. 3. In our modelling we assumed $\tau_X = 0.5$ fs (at $\hbar\omega_X = 100$ eV) and $\tau_L = 5$ fs (at $\lambda_L = 750$ nm, $\hbar\omega_L \approx 1.6$ eV), corresponding to the current state of the art. For $\tau_h < T_0/2$ (Fig. 3a–c), the energy distribution of the Auger electrons exhibits pronounced variation as the delay Δt is varied by as little as $T_0/4$ (for $\tau_h \leq T_0/5$ in close agreement with the results of the above classical treatment). From the energy distributions recorded for different values of Δt , the temporal evolution

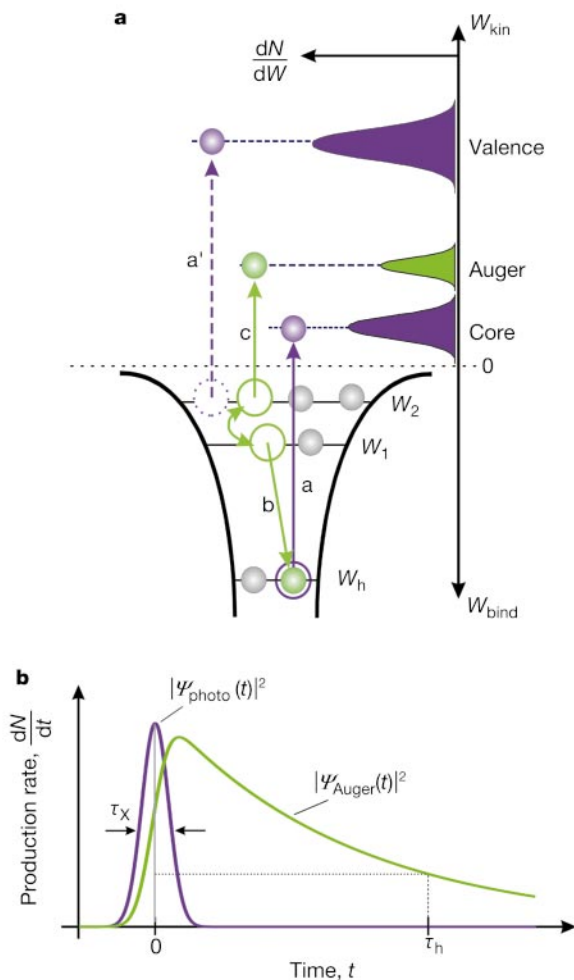


Figure 1 Schematic illustration of atomic excitation and relaxation processes following exposure to an ultrashort X-ray pulse. **a**, In the absence of resonances, the atom instantly responds by ejecting photo electrons (processes a, a') with $|\psi_{\text{photo}}(t)|^2$ following the temporal intensity profile of the exciting X-ray pulse. The inner-shell vacancy (W_h) is subsequently filled (process b) by an electron from an outer shell upon emission (process c) of a secondary (Auger) electron with $|\psi_{\text{Auger}}(t)|^2$ tracing the decay of the inner-shell vacancy. **b**, Probing the emitted photo and Auger electrons reveals the excitation and relaxation dynamics of core-excited atoms. W_{kin} , kinetic energy; W_{bind} , binding energy; W_h , $W_{1,2}$, binding energies of core and valence electrons, respectively; N , number of freed electrons. τ_X , duration of X-ray pulse; τ_h , decay time of core hole. The photo and Auger electrons are represented here (as in Fig. 5) in violet and green, respectively.

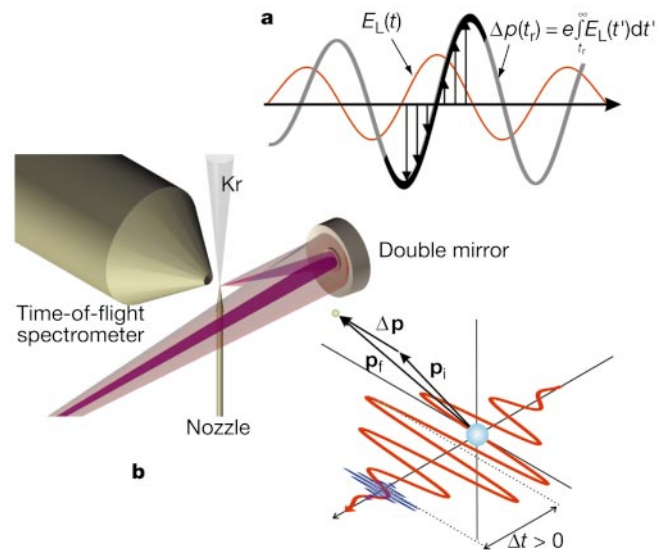


Figure 2 Attosecond two-colour sampling technique for probing electron emission from atoms. An extreme ultraviolet or X-ray pulse excites the atomic target and induces electron emission (see Fig. 1a). A delayed probe light pulse transfers a momentum Δp to the ejected electron after its release. p_i and p_f represent the electron's initial and final momentum, respectively. **a**, The transferred momentum sensitively depends on the phase and amplitude of the light field vector $E_L(t)$ at the instant of release resulting in a time-to-energy mapping on an attosecond timescale. For processes lasting less than the light cycle the oscillating light field constitutes a sub-femtosecond probe, whereas processes lasting longer than the light-wave cycle are sampled by the amplitude envelope of the laser pulse. In both cases, a sequence of light-affected electron energy spectra are recorded at different delays Δt , from which the time evolution of electron emission is reconstructed. **b**, In our experiments, we used a 97-eV, sub-femtosecond soft-X-ray pulse for excitation and a 750-nm (1.6 eV), sub-7-fs few-cycle light pulse for probing electron emission. The two pulses are collinearly focused into a krypton gas target by a two-component mirror similar to that used in ref. 5 but designed to reflect photons with higher energies. The kinetic energy distribution of the ejected photo and Auger electrons has been measured with a time-of-flight spectrometer aligned parallel to the polarization direction of the light pulse and the X-ray pulse.

of both the emission rate and a possible energy sweep of the electrons may be retrieved with a resolution of about $T_0/20$ (≈ 100 as at $\lambda_L = 750$ nm) in future studies of sub-femtosecond inner-shell dynamics.

For τ_h approaching and exceeding T_0 (Fig. 3c–e), which is relevant to our current experiment, variation of the energy distribution with Δt within the laser cycle gradually disappears. The energy spectrum is broadened and broken up into sidebands spaced by $\hbar\omega_L$ at any Δt within the range of temporal overlap of electron emission and the laser pulse^{13–16}. The sidebands originate from quantum interferences between different portions of the ejected electron wave packet experiencing the same momentum shift by the laser field. Although the overall width of the broadened energy distribution can still be estimated from the classical treatment, the emergence of the sideband structure can only be accounted for by quantum models. For electron emission times comparable to or longer than T_0 (Fig. 3d, e) the number of electrons scattered into sidebands is predicted to be proportional to the convolution of electron emission and the laser pulse. Hence for dynamics extending over T_0 or longer, evaluation of the X-ray-pump/visible-light-probe experiments does not rely on the energy–time mapping (in equation (1)) required for sub-cycle dynamics. Rather, the sideband area versus Δt directly yields the temporal evolution of electron emission (by deconvolution), provided that the laser pulse is sufficiently short (a few femtoseconds) in duration.

Recently we applied this general concept to sampling photo-electron emission with a few-cycle probe light pulse. From the measured photo-electron emission time we obtained the duration of the ionizing soft-X-ray pulse^{5,8}. The measurements revealed that few-cycle light-driven atoms are capable of emitting isolated soft-X-ray pulses ($\hbar\omega_X \approx 100$ eV) with sub-femtosecond duration ($\tau_X < 0.5$ fs) in excellent synchrony ($\Delta t_{\text{jitter}} < 0.2$ fs) with the phase of its few-cycle ($\tau_L \approx 7$ fs) driver light wave⁸. These pulses,

along with the above concept for electron sampling, combine to give an attosecond-resolution pump–probe technique for tracing electron dynamics after core-level excitation of atoms (or molecules).

Tracing the decay of an M-shell vacancy in krypton

For the first proof-of-principle experiments we selected one of the most comprehensively studied atomic inner-shell relaxation processes, the MNN Auger decay in krypton^{17–20}. The Lorentz-type line profiles of the resultant Auger electrons imply an exponential decay of the core hole with a few-femtosecond lifetime. Using the notation of Fig. 1, in this case W_h , W_1 and W_2 correspond to the $3d$, $4s$ and $4p$ sub-shells, respectively.

The pump–probe apparatus used here has been described elsewhere⁵. Briefly, the few-cycle light pulse ($\tau_L \approx 7$ fs, $\lambda_L = 750$ nm, $T_0 = 2.5$ fs) is delivered in an annular beam and hits the outer part of the two-component mirror shown in Fig. 2b. The soft-X-ray pulse with a continuum spectrum extending to photon energies of around 100 eV (ref. 8) propagates in a collinear beam matching the diameter of the inner part of the two-component mirror. For the current experiments, a dedicated Mo/Si multilayer mirror with a reflectance curve peaking at 97 eV has been manufactured. The energy spectrum of the X-ray pulse filtered by the mirror was upshifted as compared to previous experiments^{5,8} to avoid the excitation of $3d$ electrons into Rydberg states, which comes into play at $\hbar\omega_X < 95$ eV. The energy upshift comes at the expense of a narrower bandwidth of the Mo/Si multilayer (full-width at half-maximum, FWHM ≈ 3 eV), resulting in an increased pulse duration, $\tau_X \approx 0.9$ fs, as compared to recent experiments^{5,8}.

Both the X-ray and the light pulse are polarized along the axis of the time-of-flight electron energy spectrometer (Fig. 2b) to collect the electrons with a final momentum directed nearly parallel to the laser polarization. This parallel detection geometry is characterized by $\theta \approx 0^\circ$ and ‘activates’ the last term in equation (1). A comparison

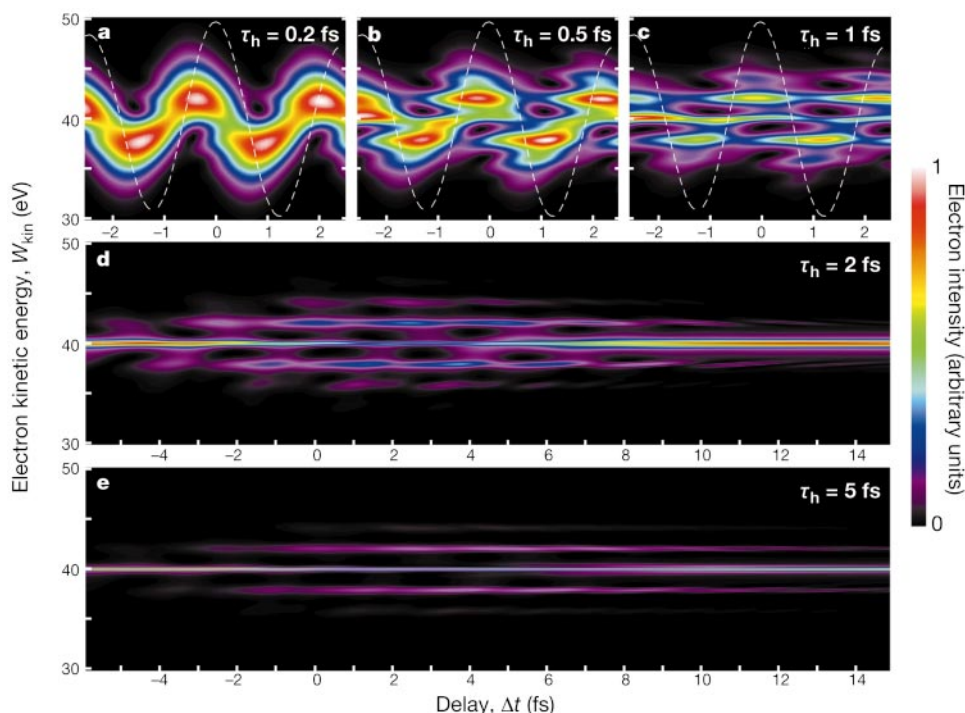


Figure 3 Laser sampling of Auger electron emission with attosecond resolution. Simulated electron spectra of a single isolated Auger line resulting from a hypothetical atomic core hole decay of different decay time constants τ_h in the presence of a few-cycle 750-nm laser field ($T_0 = 2.5$ fs) with $\varphi \approx 0$ using our quantum mechanical model (Supplementary Information). The spectra are shown in false-colour representation at

small delays near $\Delta t = 0$ for sub-femtosecond decays (a, b, c) and over an extended delay range for decays lasting longer than the wave cycle (d, e). The dashed line represents the laser electric field. The durations of the exciting X-ray and probing laser pulse and the peak intensity of the latter were assumed to be $\tau_X = 0.5$ fs, $\tau_L = 5$ fs (FWHM) and $I_0 = 5 \times 10^{11}$ W cm⁻², respectively.

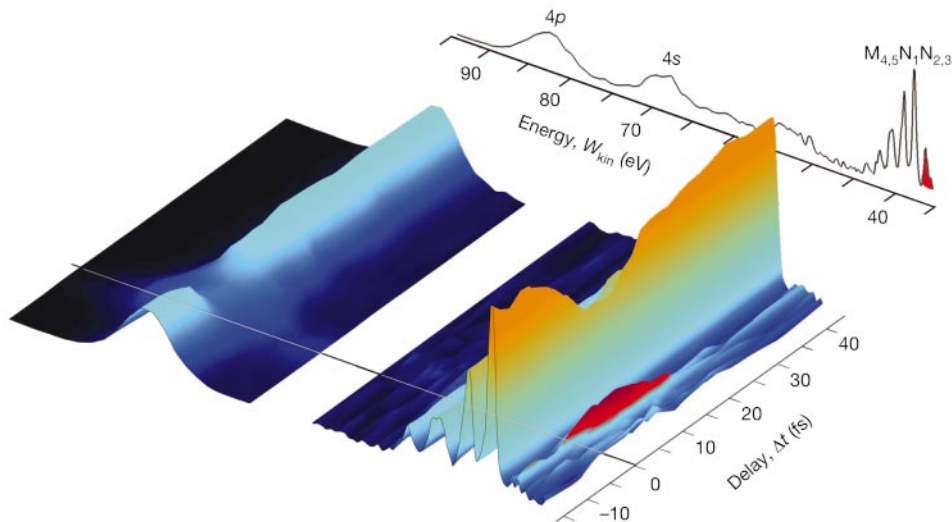


Figure 4 Evolution of electron spectra following core excitation. The spectra were recorded after the exposure of krypton atoms to a 97-eV sub-femtosecond X-ray pulse and a sub-7-fs, 750-nm laser pulse at different delays between the pulses. The surface plot shows the evolution of the 4p photo line and the $M_{4,5}N_1N_{2,3}$ Auger lines with Δt . The spectrum highlighted in the background was recorded at $\Delta t = 5$ fs. According to equation (1) the kinetic energy spectrum of the sub-laser-cycle-duration photo electrons (4s and 4p lines) should be periodically shifted versus Δt . However, for τ_X comparable to $T_0/2$ this periodic shift merges to a continuous broadening (probing the laser amplitude

envelope $E_a(t)$) owing to the absence of stabilization of the absolute phase φ in our laser pulses, because light pulses with $\varphi \approx 0$ and $\varphi \approx \pi$ induce shifts of opposite sign and turn asymmetric shifts into symmetric broadening. The laser peak intensity was evaluated from the broadening of photo lines as $I_p \approx 5 \times 10^{11} \text{ W cm}^{-2}$. The evolution of the first-order sideband (highlighted in red) of the lowest-energy member of the $M_{4,5}N_1N_{2,3}$ Auger group reflects the delayed decay dynamics of the krypton 3d core hole. A pronounced positive temporal shift of the side-band maximum with respect to the photo-line minimum is clearly discernible (see also Fig. 5).

of the second and third terms in equation (1) reveals that, for $W_i \gg \hbar\omega_L$, this results in a detectable change of the electron energy at substantially lower light-field strengths compared to the orthogonal detection geometry ($\theta \approx 90^\circ$). The low field strength helps to avoid both multiphoton ionization and a perturbation to the studied electronic relaxation. The experimental results are summarized in a surface plot compiled from 20 spectra recorded at a sequence of pump–probe delays Δt (defined in Fig. 2b), each one integrated over 300,000 laser shots and normalized to constant integral electron counts (Fig. 4). By definition, $\Delta t = 0$ at the coincidence of the peaks of the X-ray and laser pulses, which has been experimentally determined by observing the maximum broadening of the 4p and 4s photo lines.

The improved transmittivity and spectral resolution (~ 0.5 eV) of the time-of-flight spectrometer at lower kinetic energies favours the $M_{4,5}N_1N_{2,3}$ Auger group around 40 eV for an in-depth analysis. Before doing this we can make a few instructive observations. The photo lines are broadened smoothly (without the appearance of sidebands) in the range of temporal overlap between the laser and the X-ray pulse, indicating the sub- T_0 duration of the emitted electron wave packet (for more detailed comments, see legend of Fig. 4). In strong contrast, the Auger lines do not exhibit notable broadening. Instead, they are redistributed into sidebands spaced by $\hbar\omega_L \approx 1.6$ eV under the influence of the laser field. This is indicative of an Auger emission time comparable to or longer than T_0 according to our above discussion. The conspicuous suppression of the most prominent Auger peak and the simultaneous appearance of its first-order sideband (highlighted in red; other sidebands are hidden in the MNN group) indeed survive the broadening of the photo lines (for $\Delta t > 0$), confirming prolonged emission of the Auger electrons compared to that of photo-electrons.

For a quantitative analysis of the Auger emission we evaluated the sideband area A_{sb} (highlighted in red in Fig. 4) from the normalized electron spectra by combined gaussian peak-fitting to the whole $M_{4,5}N_1N_{2,3}$ group. The evolution of $A_{sb}(\Delta t)$ is shown by the circles in Fig. 5a. Figure 5b depicts the broadening $\delta w(\Delta t) = [\Delta w(\Delta t)^2 - \Delta w(\infty)^2]^{1/2}$ of the width Δw of the 4p photo line (circles) recorded

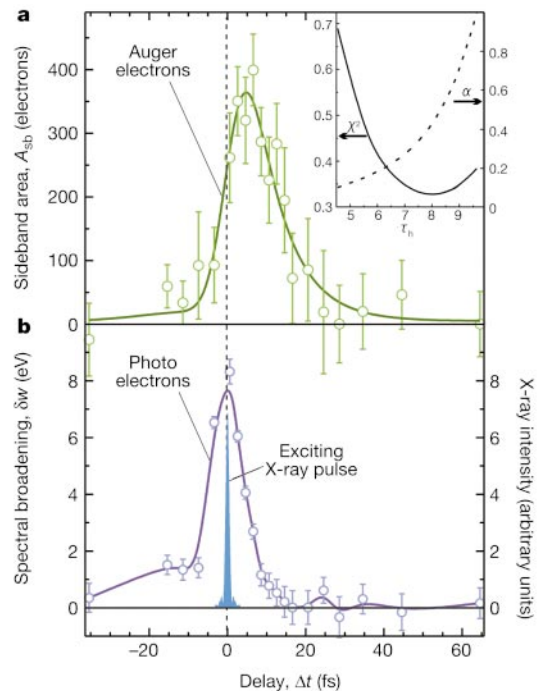


Figure 5 Probing the temporal evolution of Auger electron emission. **a**, Sideband area (circles) of the first-order sideband of the lowest-energy Kr $M_{4,5}N_1N_{2,3}$ Auger line extracted from the spectra of Fig. 4 and a fitted convolution of the exponential decay curve $R_0 \int \exp[-(t-t')/\tau_h] E_X^2(t') dt'$ with a power 2α of the amplitude envelope of the laser field. Because $\tau_X \ll \tau_L$, the latter is proportional to the spectral broadening of the 4p photoline from the same electron spectrum, presented in **b**. Panel **b** also displays the temporal intensity profile of the X-ray pulse computed from its measured energy distribution (FWHM ≈ 3 eV) assuming the absence of chirp, which applies to cut-off harmonic radiation under our experimental conditions as was verified in a recent experiment⁶. Inset, the confidence band for τ_h is confined by the merit function χ^2 and the valid range of α to $\tau_h = 7.9^{+1.0}_{-0.9}$ fs.

at the same instants. For $\tau_{el} \ll \tau_L$ and $\delta w(0)^2 \gg \Delta w(\infty)^2$, according to equation (1), the broadening displays the amplitude envelope $E_a(t)$ of the light field with the leading edge of the pulse being sampled at positive delays $\Delta t > 0$. A spline fit to the measured data (line in Fig. 5b) yields a FWHM duration of the field envelope $E_a(t)$ of slightly less than 10 fs, in excellent agreement with the result of a simultaneous autocorrelation measurement of the laser pulse ($\tau_L \approx 7$ fs). From $\delta w(\Delta t)$ we can also determine the start of the Auger emission ($\Delta t = 0$) with an accuracy of ± 0.5 fs.

In the low-intensity (perturbative) limit, the transition rate into the first-order sideband is proportional to the instantaneous laser intensity $E_a^2(t)$. However, for increasing intensities, depletion of the main peak and transition from the first-order into the second-order sideband causes deviation from this scaling. A modified scaling law of the form $E_a^{2\alpha}(t)$ with $0 < \alpha < 1$ represents a convenient approximation¹⁶. We confirmed the validity of this model for the relevant intensity range (10^{11} – 10^{12} W cm⁻²) experimentally by measuring A_{sb} as a function of the laser-pulse energy at a fixed delay Δt . We obtained $\alpha = 0.5 \pm 0.2$, in close agreement with previous findings¹⁶ about the effective sublinear scaling of A_{sb} with laser intensity. We can now obtain quantitative dynamical information about the decay of the krypton M-shell vacancies by fitting the convolution of the Auger emission rate $R_0 \int \exp[-(t-t')/\tau_h] E_X^2(t') dt'$ and the effective sampling function $E_a^{2\alpha}(t - \Delta t)$ to the measured sideband area $A_{sb}(\Delta t)$. With $\Delta t = 0$ and $E_a(t)$ known from the photoelectron data and $E_X^2(t)$ estimated (Fig. 5b) to be gaussian with $\tau_X = 0.9$ fs, the only fit parameter that remains to be determined is the lifetime of the core hole, τ_h .

To double check the validity of our effective sampling function we have computed the best fit for several values of α between 0 and 1. The inset in Fig. 5a shows the optimum value of τ_h as a function of α along with the corresponding value of the merit function χ^2 . The best-quality fit (represented by the line in Fig. 5a) has been obtained for $\alpha \approx 0.4$, which is within the above confidence interval acquired from an independent measurement. This analysis yields $\tau_h = 7.9^{+1.0}_{-0.9}$ fs for the lifetime of M($d_{5/2}$) vacancies in krypton. The measured lifetime yields an Auger linewidth of $\Gamma = \hbar/\tau_h = 84 \pm 10$ meV (not resolved by our spectrometer). This is in good accordance with 88 ± 4 meV, the result of recent energy-domain measurements¹⁹. The good agreement also indicates that post-collision interactions²¹ have little, if any, influence on our time-resolved Auger spectra in spite of the near-threshold (97 eV) excitation.

Exploring atomic dynamics by time-domain studies

The exponential decay of krypton M-shell vacancies served as a benchmark process for testing the feasibility of time-resolved atomic inner-shell spectroscopy. The measured decay time can also be inferred from energy-domain measurements.

We believe we may nevertheless expect new insight into atomic dynamics from time-domain experiments. Whenever continuum states are involved in the atomic relaxation dynamics, several interfering relaxation pathways may connect the same initial and the same final state, resulting in a complex non-exponential temporal behaviour. For example, in the photoionization process, Fano-type²² or shape²³ resonances are indicative of the outgoing electron being captured by its parent ion in a transient state for a short while. In the Auger process, post-collision interactions²¹ tend to give rise to complications. Measuring the energy sweep and (non-exponential) envelope of the emitted electron wave packets with attosecond resolution will provide comprehensive insight into the transient states involved. Full exploration of the dynamical behaviour of multi-electron systems will therefore require time-domain

spectroscopy to complement energy-domain measurements.

Atomic inner-shell processes have so far been studied only in the absence of external perturbations (other than the particle initiating the process). However, in high-field interactions atoms are often exposed to light electric-field strengths comparable to that of the effective Coulomb fields acting on inner-shell electrons. Atoms exposed to atomic-scale light fields (relevant in, for example, X-ray laser research) may exhibit a dynamical behaviour significantly different from that of unperturbed atoms. Because of the short-lived optical perturbation this can be investigated exclusively by time-domain techniques.

The evolution of time-resolved atomic physics, what might be dubbed attophysics, and its impact on other fields will—beyond experimental advances—critically depend on progress in the theoretical understanding of the dynamics of excited multi-electron systems and their interaction with strong fields. □

Received 5 August; accepted 20 September 2002; doi:10.1038/nature01143.

1. Töpler, A. Vibroskopische Beobachtungen über die Schwingungsphasen singender Flammen (der chemischen Harmonica) mit Benutzung des Schlierenapparates. *Ann. Phys. Chem.* **128**, 126–139 (1866).
2. Zewail, A. Femtochemistry: atomic-scale dynamics of the chemical bond (adapted from the Nobel Lecture). *J. Phys. Chem. A* **104**, 5660–5694 (2000).
3. Bromage, J. & Stroud, C. R. Jr Excitation of three-dimensionally localized atomic electron wave packet. *Phys. Rev. Lett.* **83**, 4963–4966 (1999).
4. Drescher, M. et al. X-ray pulses approaching the attosecond frontier. *Science* **291**, 1923–1927 (2001) published online 15 February 2001; 10.1126/science.1058561.
5. Hentschel, M. et al. Attosecond metrology. *Nature* **414**, 509–513 (2001).
6. Papadogiannis, N. A., Witzel, B., Kalpouzos, C. & Charalambidis, D. Observation of attosecond light localization in higher order harmonic generation. *Phys. Rev. Lett.* **83**, 4289–4292 (1999).
7. Paul, P. et al. Observation of a train of attosecond pulses from high harmonic generation. *Science* **292**, 1689–1692 (2001).
8. Kienberger, R. et al. Steering attosecond electron wave packets with light. *Science* **297**, 1144–1148 (2002) published online 11 July 2002; 10.1126/science.1073866.
9. Krausz, F. Attosecond spectroscopy comes of age. *Opt. Photon. News* **13**, 62–68 (May 2002).
10. Krause, M. O. Atomic radiative and radiationless yields for K and L shells. *J. Phys. Chem. Ref. Data* **8**, 307–328 (1979).
11. Itatani, J. et al. Attosecond streak camera. *Phys. Rev. Lett.* **88**, 173903 (2002).
12. Kitzler, M., Milosevic, N., Scrinzi, A., Krausz, F. & Brabec, T. Quantum theory of attosecond XUV pulse measurement by laser-dressed photoionization. *Phys. Rev. Lett.* **88**, 173904 (2002).
13. Freeman, R. R. & Bucksbaum, P. H. Investigation of above-threshold ionization using subpicosecond pulses. *J. Phys. B* **24**, 325–347 (1991).
14. Schins, J. M. et al. Observation of laser-assisted Auger decay in argon. *Phys. Rev. Lett.* **73**, 2180–2183 (1994).
15. Glover, T. E., Schoenlein, R. W., Chin, A. H. & Shank, C. V. Observation of laser assisted photoelectric effect and femtosecond high order harmonic radiation. *Phys. Rev. Lett.* **76**, 2468–2471 (1996).
16. Bouhal, A. et al. Cross-correlation measurement of femtosecond nonlinear high-order harmonics. *J. Opt. Soc. Am. B* **14**, 950–956 (1997).
17. Aksela, H., Aksela, S. & Pulkkinen, H. Correlation effects in $4s^0 4p^6$ and $4s^1 4p^5$ configurations of krypton studied by the M-NN Auger decay. *Phys. Rev. A* **30**, 2456–2461 (1984).
18. Carlson, T. A. et al. Angular distribution of ejected electrons in resonant Auger processes of Ar, Kr, and Xe. *Phys. Rev. A* **39**, 1170–1185 (1988).
19. Jurvansuu, M., Kivimäki, A. & Aksela, S. Inherent lifetime widths of Ar $2p^{-1}$, Kr $3d^{-1}$, Xe $3d^{-1}$, and Xe $4d^{-1}$ states. *Phys. Rev. A* **64**, 012502 (2001).
20. Schmidtke, B. et al. The Kr $M_{4,5}N_1N_{2,3}$ (1P_1) Auger decay: measurement of the transferred spin polarization and analysis of Auger amplitudes. *J. Phys. B* **34**, 4293–4310 (2001).
21. Borst, M. & Schmidt, V. Vanishing postcollision interaction in inner-shell photoionization. *Phys. Rev. A* **33**, 4456–4458 (1986).
22. Fano, U. Effects of configuration interaction on intensities and phase shifts. *Phys. Rev.* **124**, 1866–1878 (1961).
23. Becker, U. et al. Subshell photoionization of Xe between 40 and 1000 eV. *Phys. Rev. A* **39**, 3902–3911 (1989).

Supplementary Information accompanies the paper on Nature's website (<http://www.nature.com/nature>).

Acknowledgements Sponsored by the Fonds zur Förderung der wissenschaftlichen Forschung (Austria), the Deutsche Forschungsgemeinschaft (Germany) and by the European Atto Network.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to M.D. (e-mail: drescher@physik.uni-bielefeld.de) or F.K. (e-mail: ferenc.krausz@tuwien.ac.at).

Stochastic and genetic factors influence tissue-specific decline in ageing *C. elegans*

Laura A. Herndon*, Peter J. Schmeissner*, Justyna M. Dudaronek*, Paula A. Brown*, Kristin M. Listner*, Yuko Sakano*, Marie C. Paupard†, David H. Hall† & Monica Driscoll*

* Department of Molecular Biology and Biochemistry, Rutgers, The State University of New Jersey, A232 Nelson Biological Laboratories, 604 Allison Road, Piscataway, New Jersey 08854, USA

† Department of Neuroscience, Albert Einstein College of Medicine, 1410 Pelham Parkway, Bronx, New York 10461, USA

The nematode *Caenorhabditis elegans* is an important model for studying the genetics of ageing, with over 50 life-extension mutations known so far. However, little is known about the pathobiology of ageing in this species, limiting attempts to connect genotype with senescent phenotype. Using ultrastructural analysis and visualization of specific cell types with green fluorescent protein, we examined cell integrity in different tissues as the animal ages. We report remarkable preservation of the nervous system, even in advanced old age, in contrast to a gradual, progressive deterioration of muscle, resembling human sarcopenia. The *age-1(hx546)* mutation, which extends lifespan by 60–100%, delayed some, but not all, cellular biomarkers of ageing. Strikingly, we found strong evidence that stochastic as well as genetic factors are significant in *C. elegans* ageing, with extensive variability both among same-age animals and between cells of the same type within individuals.

In *C. elegans*, many single-gene mutations confer significant extensions of lifespan¹. Life-prolonging mutations affect several aspects of nematode biology, including insulin-like signalling for dauer larvae development, the development and function of sensory neurons, the consumption of food, and biochemical defences against stresses such as heat shock and reactive oxygen species (reviewed in refs 2, 3). However, a detailed understanding of the molecular and cellular events influencing longevity remains a central mystery in the biology of ageing. Similarly, factors that regulate healthspan (the adult period of unimpaired activity and function that precedes age-related decline) have not been well defined, despite their potential as targets for forestalling the deleterious consequences of ageing.

A major gap in our understanding of *C. elegans* ageing is the question of what happens to cell and organ systems over time. The paucity of information regarding the cellular consequences of ageing is striking, given that each of the 959 somatic cells in the mature *C. elegans* hermaphrodite is an identified cell of known ancestry⁴, which can be readily visualized in the living animal by using reporters labelled with green fluorescent protein (GFP), and that the reconstruction of full-animal serial section electron micrographs has documented the subcellular architecture of most cells⁵. To facilitate the elaboration of ageing mechanisms and to extend our understanding of tissue-specific contributions to healthspan and longevity, we characterized major cellular changes that accompany *C. elegans* ageing by using fluorescence and electron microscopic techniques.

A stochastic component to age-related locomotory decline

With abundant food at 20 °C, *C. elegans* develops from an embryo to sexual maturity in approximately 3 days, lays virtually all of its eggs over the first 3–4 days of adulthood, and then persists through a post-reproductive period during which senescent decline is evident (mean lifespan approximately 12–18 days)⁶. To lay the groundwork for a detailed analysis of cellular changes that comprise ageing in *C. elegans*, we first examined the appearance of wild-type animals over time using Nomarski differential-interference-contrast optical microscopy. In addition to an accumulation of dark pigment and

lipofuscin within the body^{6,7}, vacuole-like structures, visually similar to neurons that die in response to ion channel injury⁸, first appear in nematode bodies towards the end of the reproductive period, and increase in size and number over time (Supplementary Fig. 1a). These vacuolar structures occur in variable positions and with a variable time of onset.

Our initial look at ageing populations also suggested significant heterogeneity in mobility of same-age individuals; we therefore

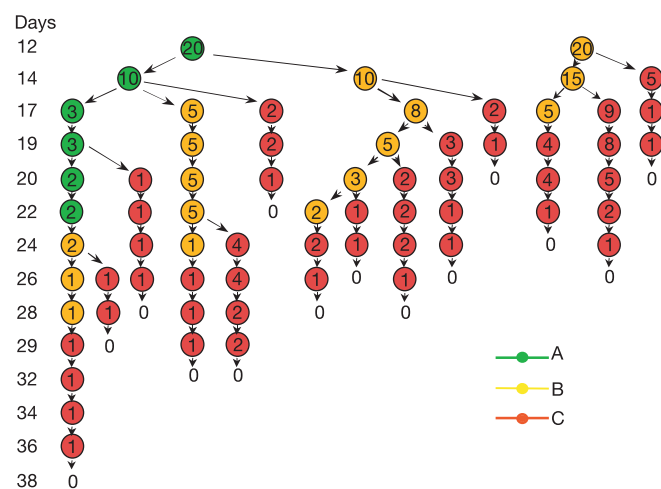


Figure 1 Behavioural phenotypes of ageing nematodes indicate progressive decline with stochastic onset and variable rates. Twelve days after a synchronized egg lay, we categorized 40 wild-type animals on the basis of their locomotory phenotypes A (green), B (yellow) and C (red). We reclassified individuals every 1–3 days until all animals had died. Numbers within the circles represent animals scored in a given class on a particular day. Similar results were observed in four additional independent trials. Of 217 class A animals, 194 proceeded through stages B and then C before death. Similarly, 254 of 262 class B animals became class C animals before dying. We never observed a reversal of class order (B to A or C to B).

characterized locomotion in ageing populations in detail. We scored age-synchronized individuals both for spontaneous movement and for response to prodding with a wire at various points during adulthood. Somewhat unexpectedly for an isogenic population, we could distinguish three classes of behavioural phenotypes in a same-age population late in the lifespan. In one class, animals move constantly and leave sinusoidal tracks in the *Escherichia coli* lawn through which they migrate. These highly mobile animals, which we designate as class A, respond to prodding by vigorous movement away from the touch stimulus. Members of class B do not move unless prodded and leave tracks that are non-sinusoidal, reflecting uncoordinated locomotion. Class C animals do not move forward or backward even when prodded but do exhibit head and/or tail movements or twitch in response to touch. These classifications are broad enough for the vast majority of nematodes to be fitted into a single class after one observation. Although borderline cases did arise, class was typically clear by the next score point. All animals begin adulthood in class A. Class B animals first appear in the population at about day 6–7, which coincides approximately with the end of the reproductive period under laboratory conditions at 20 °C. Class C animals first appear at about day 9–10 (Supplementary Fig. 1b).

To gain a better understanding of the origin of the diversity of age-related behavioural classes, we followed hundreds of individuals over time. We sorted age-synchronized animals by class and re-scored for class every 1–3 days until death (Fig. 1 shows data from one representative trial). This analysis confirmed that age-associated locomotory defects increase progressively in severity over time^{7,9}, and established that age-related behavioural changes occur with different times of onset and progress at different rates of decline in individual animals. We found that behavioural class is a

better predictor of life expectancy than chronological age (Supplementary Fig. 1c).

That both the individual time of onset and the rate of behavioural decline are widely variable among isogenic members of a same-age population reared under the same environmental conditions indicates that at least one major factor in the behavioural ageing of *C. elegans* is stochastic.

Intactness of the nervous system in old nematodes

We initially sought evidence of age-related neurodegeneration in *C. elegans* because diminished locomotory responses could result from neuronal structural deterioration or cell death, and because the vacuolar structures in ageing nematodes superficially resemble necrotic-like neuronal cell deaths¹⁰. We used GFP reporters to visualize specific groups of neurons from comparably aged, or comparably categorized (locomotory classes A, B or C), nematodes over the lifespan.

Examination of animals expressing an integrated promoter for the *unc-119* gene that expresses GFP (*p_{unc-119}GFP*) array in all neurons¹¹ (3–4-day-olds ($n = 14$), 6–9-day-olds ($n = 29$) and 12–16-day-olds ($n = 15$)) revealed that all major structures of the *C. elegans* nervous system, including dorsal and ventral nerve cords, neuronal commissures that cross the body, lateral neurons and the nerve ring⁵, remain clearly distinguishable throughout the lifespan. Even in 15 late-stage class C animals ($n = 7$ 14-day-olds; $n = 8$ 18-day-olds), neuronal processes were positioned and extended similarly to those in young animals (data not shown). For a more detailed analysis of individual neurons, we scored animals harbouring an integrated *p_{mec-4}GFP* array expressed in six body touch-sensing neurons¹² and found that, without exception, we could easily distinguish normal cell bodies and processes at all

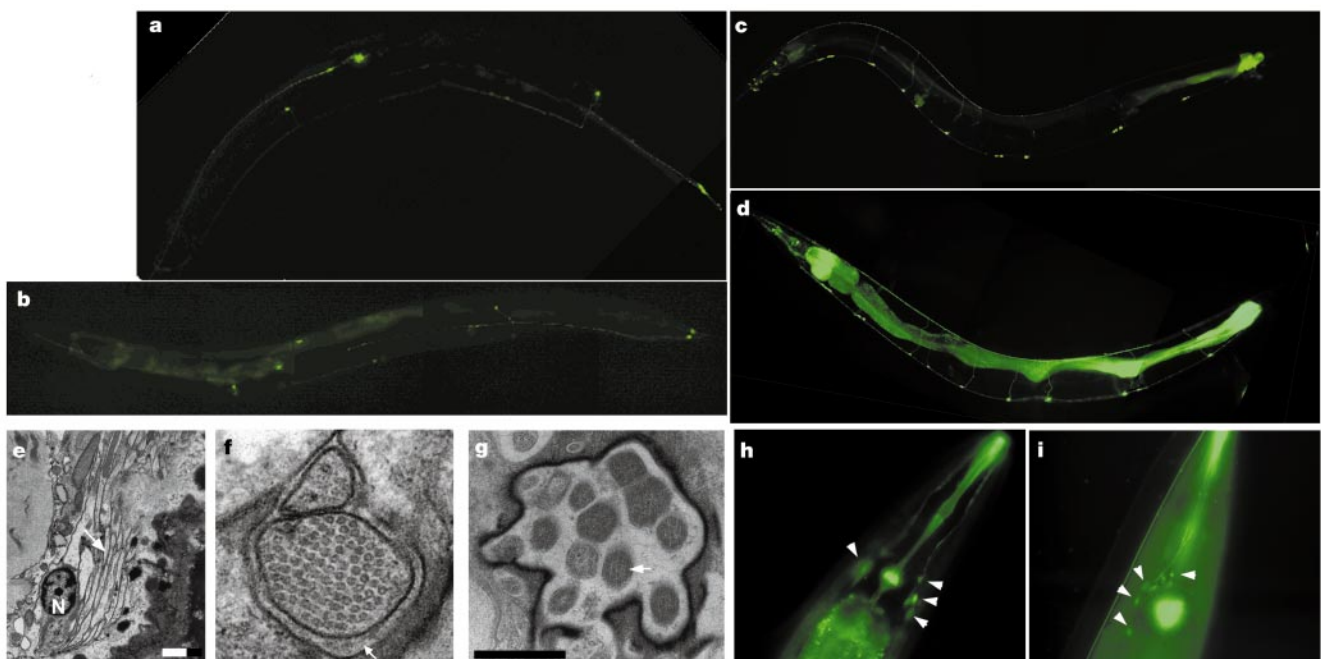


Figure 2 The integrity of the nervous system is maintained in ageing nematodes. **a, b**, Six touch neurons were revealed with an integrated *p_{mec-4}GFP* reporter; examples of young (**a**, day 4) and old (**b**, day 15) adults are shown. **c, d**, Eleven DA and DB motor neurons were revealed with a *p_{unc-129}GFP* reporter; examples of young (**c**, day 4) and old (**d**, day 15) animals are shown. At every time point, the 11 neurons scored could be identified and processes appeared intact (day 4, $n = 77$ worm/847 cell observations; day 15, $n = 82$ worm/902 cell observations). Background intestinal autofluorescence is also apparent and increases with age. **e**, Section through a day 18 class C animal reveals processes in the nerve cord (white arrow) and a neuronal nucleus (N) that are

ultrastructurally similar to those in younger animals. Scale bar, 1 μ m. **f**, Touch neurons maintain characteristic features such as prominent 15-protofilament microtubules (circular in cross-section) and an extracellular mantle that borders the neuron process (white arrow), even in a day 18 class A animal. Scale bar, 0.1 μ m. **g**, Electron micrograph of an amphid cross-section reveals ciliated dendrites of all 12 amphid neurons (arrow indicates one of these) in a class B animal. Scale bar, 0.5 μ m. **h, i**, A subset of amphid sensory neurons are open to the environment, and neurons fill with dye in both young (**h**, day 4) and old (**i**, day 24, class C) animals. Arrowheads indicate filled cell bodies.

timepoints ($n = 180$ animals; 1,080 total cell observations; Fig. 2a, b). In 20 decrepit C-stage animals, we found six touch neurons intact in every animal (120 total cell observations). We also used an integrated $p_{unc-129}GFP^{13}$ reporter to examine the number and integrity of the DA and DB motor neurons (which innervate body wall muscle) and found all scored neurons detectable with commissures intact (2,453 total neurons evaluated; $n = 77$ 4–6-day-olds; $n = 64$ 8–10-day-olds; $n = 82$ 12–17-day-olds; Fig. 2c, d). Similarly, 14 class C animals of various chronological ages ($n = 7$ 15-day-olds; $n = 5$ 19-day-olds; $n = 2$ 30-day-olds) had DA and DB motor neurons indistinguishable from DA and DB neurons in

younger adults (data not shown). In sum, a detailed examination of representative sensory and motor neurons failed to indicate any signs of neuronal degeneration associated with either advanced *C. elegans* age or extensive behavioural decline.

We also examined thin-section electron micrographs derived from 18-day-old nematodes of behavioural classes A, B and C and compared them with the extensive collection of electron micrographs of young adult animals available at the Center for *C. elegans* Anatomy. This survey identified intact neurons even in class C animals and failed to produce any evidence of widespread degeneration in the nervous system. The nerve rings and longitudinal nerves

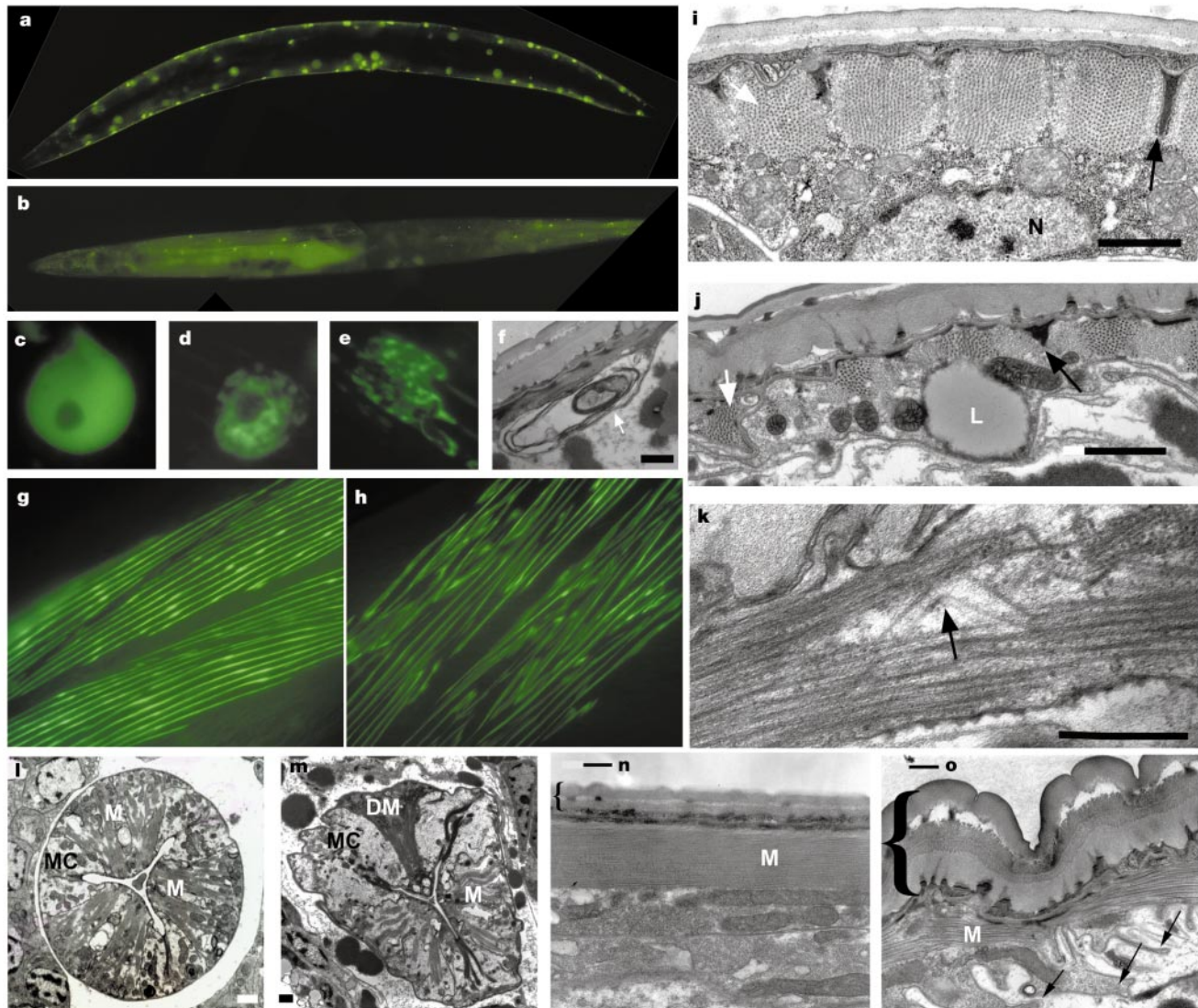


Figure 3 Age-related deterioration of *C. elegans* body wall muscle. **a, b**, Whole-animal view of nematodes expressing GFP localized to the nuclei of body wall muscle (**a**, day 8; **b**, day 14; anterior to the left). The apparent disappearance of signal seen at $\times 50$ magnification actually reflects the conversion of nuclear GFP to a fragmented pattern (see **e**). **c–e**, Magnified ($\times 1,000$) view of muscle nuclei as revealed by GFP (**c**, day 7; **d**, day 12, appearance of small, dark patches and an increase of size of the nucleolus; **e**, day 18, fragmented phase). **f**, Example of a rare muscle nucleus (arrow) undergoing autophagy; day 18, class B. Scale bar, $1\ \mu\text{m}$. **g, h**, Body wall muscle sarcomeres as detected by a $p_{myo-3}MYO-3/GFP$ translational fusion, highlighting myosin heavy chain A (**g**, day 4; **h**, day 18, class B). **i–k**, Electron micrographs of body wall muscle, scales as indicated. **i**, Cross-section through a single muscle quadrant of a day 4 animal. Sarcomeres (white arrow) interface with the hypodermis and cuticle via dense bodies (black arrow). Scale bar, $1\ \mu\text{m}$. **j**, Cross-section through single muscle of a day 18 class B animal with prominent

lipid inclusion (L). Note the loss of cytoplasmic volume and loss of myosin thick filaments (white arrow), but a retention of actin and dense bodies that interface with the hypodermis and cuticle (black arrow). Scale bar, $1\ \mu\text{m}$. **k**, Magnified view of a frayed sarcomere in which myosin filaments are severely bent (arrow), from a day 18 class B animal. Scale bar, $0.5\ \mu\text{m}$. **l**, Cross-section of a circular pharynx from a young adult (day 4) showing three-fold symmetry of muscle (M), marginal cells (MC) and open lumen. Scale bar, $1\ \mu\text{m}$. **m**, Cross-section of the pharynx of a day 18 class B adult (MC, marginal cell; M, muscle; DM, deteriorated muscle). Scale bar (lower left), $1\ \mu\text{m}$. **n, o**, Comparison of cuticle in young (day 4 class A, **n**) and day 18 class B (**o**) animals. Panels are at the same scale and thus the cuticle (marked by curly brackets) clearly shows differences in thickness. Disorganization of underlying muscle (M) and shedding of cytoplasm (cell fragmentation) through thin basal extensions from muscle (arrows) is also apparent in **o**. Scale bars, $0.5\ \mu\text{m}$.

of the ventral and dorsal nerve cords were well formed (Fig. 2e), and touch receptor neurons were filled with characteristic 15-protofilament microtubules and surrounded by specialized extracellular matrix typical of younger neurons (Fig. 2f). We could readily identify all 12 amphid sensory neurons of the nose (some of which have been implicated in lifespan regulation¹⁴ (Fig. 2g)), and the ciliated amphid neurons were morphologically normal. Even in 18 decrepit class C animals, the chemosensory amphid neurons that open to the environment fill with dyes comparably to their younger counterparts¹⁵ (Fig. 2h, i). Microtubule bundles and neuronal nuclear structures were generally maintained. The only subtle changes we noted were a slight increase in cytoplasmic electron density and a modest reduction in neurite calibre in older animals (slightly narrower in class C than in class B; data not shown).

Overall, our study of the ageing *C. elegans* nervous system indicates, first, that there is little if any neuronal cell death during the *C. elegans* lifespan, even in severely compromised animals; and second, that there is no large-scale age-associated loss or restructuring of neurites or nuclei. Although we cannot rule out significant changes in biophysical properties (nerve conduction, synaptic

transmission), it is clear that at the cellular level the nervous system is well maintained in senescing *C. elegans*.

Sarcopenia in ageing nematodes

Progressive locomotory impairment during *C. elegans* ageing could be the consequence of a decline in muscle function. We screened for evidence of age-related muscle deterioration by monitoring GFP-tagged proteins localized to either body wall muscle nuclei (p_{myo-3} -GFP/NLS) or sarcomeres (p_{myo-3} -MYO-3/GFP). We noted an apparent decrease, beginning in mid-life, in the number of GFP-labelled nuclei detectable under low-power magnification ($\times 50$), with class C animals exhibiting the most extreme signal loss (Fig. 3a, b). A higher-resolution examination ($\times 1,000$) indicated that this 'loss' reflects a striking change in the distribution of the nuclear-localized GFP fluorescence over time (Fig. 3c–e). Starting at day 7, small, dark patches begin to appear within many muscle nuclei (Fig. 3c). Also at day 7, the nucleus and nucleolus begin to lose circularity and become more misshapen. Another trend indicated by our nuclear GFP survey and supported by electron-microscopic data is that nucleolar size seems to increase relative to nuclear size with age (in day 7 animals, the ratios of nucleolar size to nuclear size as quantified by NIH Image were 0.16 ± 0.01 (mean $\pm$ s.e.m.) but at day 18 they were 0.30 ± 0.02 (mean $\pm$ s.e.m.)) (Fig. 3c–e; Supplementary Fig. 2a). The degree of GFP mottling increases progressively over the lifespan (Fig. 3d), so that later in life GFP fluorescence can acquire a highly fragmented appearance (Fig. 3e). The degree of nuclear change seems to correlate with locomotory class (Supplementary Fig. 2b) and to reflect changes in nuclear GFP distribution rather than nuclear breakdown, because ultrastructural electron-microscopic examination revealed nuclear breakdown only rarely (less than 2% of all muscle cells examined; Fig. 3f). Significantly, at most mid-life and late-life time points (days 9–20), we found that nuclear GFP distribution in different muscle cells within individual animals was variable (for example, ranging from non-mottled to severely mottled), indicating that stochastic factors might influence age-related nuclear changes in muscle on a cell-by-cell basis (Supplementary Fig. 2a).

Our analysis of a GFP-tagged MYO-3 protein¹⁶ that is localized to the sarcomeres of body wall muscle revealed that, whereas in young animals sarcomeres are organized in tight parallel symmetric rows (Fig. 3g), sarcomeres in older animals seem progressively disorganized with less dense packing and irregular orientation (Fig. 3h). In some individual muscles we noted marked changes in sarcomere orientation or a fraying of individual sarcomeres into thinner, misdirected strands. However, the MYO-3/GFP protein labels sarcomeres until the end of the lifespan, and thus muscles do not fully disintegrate during *C. elegans* ageing.

Age-related decline in sarcomere integrity was more marked as viewed by electron microscopy. Sarcomeres clearly become more disorganized in old muscle (Fig. 3i, j) and include significantly fewer myosin thick filaments per sarcomere unit. Within some muscles of day 18 animals, individual thick filaments appear to bend abnormally and break (Fig. 3k). We also found a striking and prevalent shrinkage of the muscle cells, due in large part to a progressive cytoplasmic loss. The plasma membrane becomes highly invaginated and fragments can be shed from the muscle belly (Fig. 3i, j; also apparent in Fig. 3n, o). We noted cytoplasmic loss even in class A animals, with loss more pronounced in classes B and C. Large lipid droplets accumulate within the muscle of aged animals (Fig. 3j). By contrast, dense bodies in muscle (attachment sites analogous to vertebrate focal adhesion sites that serve as the physical link between myofibrils and muscle membrane, the hypodermis and cuticle) are generally well maintained (Fig. 3i, j; Supplementary Fig. 3).

Muscle deterioration in ageing animals is not limited to body wall muscle. With increasing age the pharynx loses its smooth, rounded shape and takes on an irregular appearance. Myofibrils in individual pharyngeal muscles are lost over time. In addition, entire pharynx-

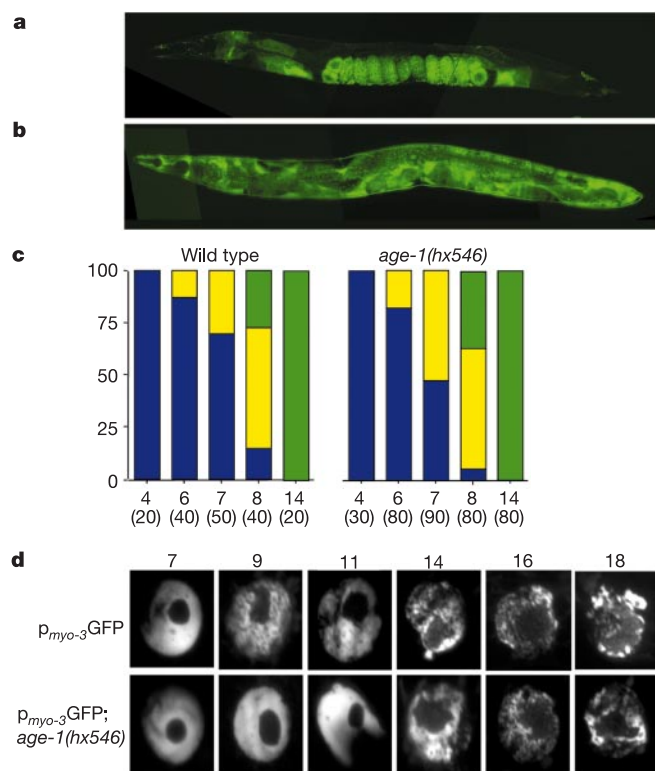


Figure 4 The long-lived *age-1* mutant is not altered in age-related yolk protein redistribution but does exhibit delayed sarcopenia. **a, b**, Animals expressing tagged yolk protein, YP170/GFP: **a**, day 4 adult with yolk localized to developing embryos and intestine; **b**, day 12 post-reproductive adult with yolk distributed throughout the body. **c**, Graph depicting scores of the yolk redistribution phenotype as determined with a YP170/GFP reporter protein; y-axis, percentage of animals in a given category; x-axis, days (total number scored). Blue represents localization only to oocytes; yellow represents localization to oocytes and body cavity; green represents localization to body cavity only. **d**, Representative muscle nuclei of the wild type and the *age-1(hx546)* mutant expressing nuclearly targeted p_{myo-3} -GFP. Although nuclei were chosen as average for their class, note that nuclear GFP changes with stochastic onset and a fair degree of variation in severity. Larger data montages used in blind tests and test results are provided in Supplementary Fig. 2. Note that in the *age-1(hx546)* background, there is a delay in the onset and progression of nuclear changes (days 7, 9 and 11), but differences are no longer clearly distinguishable after day 14.

geal muscle sectors can sporadically appear darkened and abnormal, a phenomenon that suggests a stochastic component to muscle degeneration (Fig. 3l, m).

Taken together, our results indicate that, like humans, ageing nematodes suffer from sarcopenia, the progressive loss of muscle mass and muscle function over time with mid-life onset.

Stochastic origins of catastrophic events in other tissues

Electron-microscopic data indicate that many additional cell types (such as hypodermis and intestine), but not all (the kidney-like canal cell and some oocytes), exhibit age-related deterioration (see Supplementary Fig. 4). A decline in these cell types also seems to be stochastic, much as muscle decline does. Overall, our survey indicates that ageing affects different cell types at different rates in *C. elegans* and that stochastic events are factors in the age-related decline of multiple tissue types. It might be noteworthy that hypodermis and intestine often suffer random but extreme local crises, apparently associated with plasma membrane disruption, which might contribute to the death of the animal in advanced age (see Supplementary Information Discussion on causes of death in *C. elegans* and roles of cell death in ageing).

Unregulated biosynthesis in post-reproductive *C. elegans*

Our GFP reporter/electron-microscopic analysis yielded several examples indicative of unregulated biosynthesis late in life. For example, we and others¹⁷ have noted marked accumulation of lipid inclusions in intestine (Supplementary Figs 4c, d and 5b), muscle (Fig. 3j) and hypodermis (Supplementary Fig. 5c–e). We were also struck by a marked thickening of the *C. elegans* cuticle during ageing (Fig. 3n, o). Increases in cuticle thickness can reach 10-fold that in young adults in some locales, indicating that hypodermal cells continue collagen synthesis well into adulthood, even though ageing animals have passed larval developmental moulting periods. (The cuticle of old animals is quite wrinkled, which might reflect the weakening of hypodermis and/or muscle.) Electron-microscopic data also revealed the accumulation of body-wide electron-dense material similar to that previously identified as yolk protein¹⁸. We used immunoelectron microscopy to confirm that yolk protein YP170 was distributed in electron-dense lipid-like droplets and large dark accumulations throughout the body cavity in older animals (Supplementary Fig. 5f). In addition, we could discern age-related yolk redistribution with the reporter YP170/GFP. In young animals, YP170/GFP is found almost exclusively in the intestine (where it is synthesized) as well as in late-stage oocytes (which accumulate it) and embryos (Fig. 4a), but as animals enter the post-reproductive period, we found YP170/GFP throughout the body cavity of the animal (Fig. 4b). Thus, yolk protein production continues (at least initially) in the absence of oogenesis in post-reproductive nematodes. Overall, the extensive accumulation of macromolecules, such as yolk proteins, cuticle proteins and lipid, suggest that in post-reproductive animals, a life-stage that has been subject to little (if any) natural selection pressure over time, biosynthesis/protein turnover is not tightly regulated like it is in developmental and reproductive phases.

Longevity mutations delay onset of some biomarkers of ageing

Having identified several potential biomarkers associated with ageing *C. elegans* tissues, we initiated efforts to address the basic question of how mutations causing increased longevity affect tissue-specific aspects of ageing: do long-lived mutants exhibit delay, or suppression, of all biomarkers associated with ageing, or might only specific aspects of body/tissue decline be affected? To gain first insight into this question, we compared life-long expression of the two most consistent and markedly altered GFP reporter proteins that we identified, yolk protein YP170/GFP and the muscle nuclear reporter expressed from *p_{myo-3}*GFP/NLS, in wild-type and *age-1(hx546)* mutant backgrounds. *age-1* encodes a phosphatidyl-

inositol-3-OH kinase (PI(3) kinase) that acts downstream in the DAF-2 insulin-like receptor signalling to influence lifespan¹⁹. *age-1(hx546)* animals live 60–100% longer than wild type²⁰. We categorized YP170/GFP distribution in ageing animals into three groups: fluorescence exclusively in eggs, fluorescence in eggs and body cavity, and fluorescence solely in the body cavity. When we compared wild-type and *age-1* backgrounds, we could find no significant differences in either the time of onset or the apparent extent of body-wide yolk diffusion over time (Fig. 4c), indicating that *age-1(hx546)* animals experience extended lifespan despite the fact that yolk protein becomes widely distributed throughout the body at the start of the post-reproductive period, as occurs in ageing wild-type animals. Thus, not all phenotypes associated with ageing are altered by the increased longevity mutation *age-1(hx546)*.

By contrast, when we examined changes in GFP distribution in muscle nuclei in the *age-1(hx546)* mutant background, we found that there was a significant delay in the onset and extent of nuclear GFP changes in comparison with wild-type animals (Fig. 4d). We reared *p_{myo-3}*GFP/NLS and *age-1*; *p_{myo-3}*GFP/NLS strains under identical conditions and collected photographs of every nucleus (to avoid sampling bias) from multiple nematodes from days 6 to 19. Representative nuclei are featured in Fig. 4d, but in view of the stochastic component to onset and extent of nuclear GFP redistribution we found it essential to compare large collections of nuclear photographs (Supplementary Figs 2a, c and d). Our analysis of more than 20 photographs for each time point indicated that at day 5 (young adulthood) there are no detectable differences between wild-type and *age-1* animals. By day 7, we see the onset of GFP patching in the wild type, but this is not evident in *age-1* animals. At day 9, we can find patching in both populations of nuclei but patching is clearly more extensive in the wild-type background. Differences persist until day 14/16, after which we can no longer readily distinguish between wild-type and *age-1* animals for this age-associated phenotype.

We conclude that *age-1(hx546)*, which enhances locomotory activity in ageing populations²¹, delays the onset of age-related muscle nuclear changes and suggest that the *age-1* PI(3) kinase might exert its effects on healthspan and longevity in part by prolonging muscle integrity.

A strong stochastic component for age-related decline

Our analysis of cellular changes in ageing *C. elegans* revealed several remarkable aspects of nematode senescence. Key findings are that stochastic factors make a significant contribution to senescent decline and that different cell types deteriorate at markedly different rates, with the nervous system being spared extensive cellular decline while muscle undergoes a profound and progressive deterioration with mid-life onset. We also report that the increased longevity mutation *age-1(hx546)* can delay the onset of some, but not all, cellular features of ageing.

One striking aspect of the biology of the ageing of *C. elegans* is the wide variability in both the time of onset and the rate of apparent deterioration within an isogenic population reared under uniform environmental conditions^{9,22}. Although factors in the micro-environment or life histories of individuals (for example, the amount of time spent in food as opposed to near it) could profoundly affect ageing rates, we repeatedly observed a stochastic occurrence of cellular demise within the same cell types of individual animals. The marked variation in cellular decline is probably attributed to random damage or failures. In providing evidence for significant stochastic influences on both the overall senescent decline of the organism and the degeneration of individual cells within a single *C. elegans* (an organism in which there is virtually no genetic or developmental variation between animals), our data support a key premise of the ‘disposable soma’ theory of ageing, which postulates that damage caused by chance events is a major contributor to ageing soma^{23–25}. Another premise of this theory is

that the genes that influence ageing and lifespan should primarily be involved in maintenance and repair of the soma^{23–25}, and all characterized *C. elegans* increased-longevity mutations tested increase resistance to various stresses^{3,26}. Given that stochastic contributions also affect ageing in other animals^{27,28}, we emphasize that an understanding of stochastic components will be required for a full description of ageing in both *C. elegans* and higher organisms^{27,29}.

Maintenance of nervous-system integrity in animal ageing

After scoring thousands of individual *C. elegans* neurons during ageing, we conclude that neurons maintain their structural integrity for the duration of the lifespan. At the cellular level, we find no evidence of cell loss or axon degeneration, even in animals with severely compromised mobility. Previous analyses of ageing in other nematodes also failed to note significant neuronal deterioration with age¹⁷. Similarly, neuronal counts in fly eye³⁰ and antenna³¹ provided no evidence for neuronal loss as a significant component of *Drosophila* ageing. Recent investigations using stereological imaging for quantification have generally concluded there is little neuronal loss associated with normal ageing in specific human brain regions studied, challenging long-held notions about significant loss of mammalian brain neurons with age (reviewed in ref. 32). The basic maintenance of the cellular integrity of nervous systems might be a common feature of metazoan ageing, although given documented differences in neuronal integrity among various ageing rodent inbred strains³³, it is clear that genetic factors can influence neuronal health and function in aged animals.

Because extensive neuronal loss seems less of a significant factor in age-related cognitive and motor deficits in humans, experimental attention has turned to the evaluation of synaptic function in higher organisms. A higher-resolution examination of synaptic function in nematodes (including quantification of synaptic proteins in synapses, characterization of neuronal transcripts, and detailed ultrastructural analysis of synaptic integrity in specific behavioural classes over time) will be required to identify molecular and subcellular changes that accompany neuronal ageing in *C. elegans*. Similarly, detailed changes at the neuromuscular junction remain to be examined (see Supplementary Information Discussion).

Like humans, *C. elegans* experience sarcopenia

C. elegans body wall muscle is considered to be analogous to human skeletal muscle in several respects³⁴. In humans, loss of muscle mass is associated with a shrinkage in the sizes of the fused muscle cells, thought to be due to a decrease in the cytoplasmic volume of the cell³⁵. Our electron-microscopic analysis provides clear evidence of cytoplasmic volume loss and a gross decrease in myofilament numbers in body wall muscle in aged animals. In rats, sarcomere integrity is progressively compromised with age such that the packing geometry of the sarcomere fibres is generally loosened, and fraying and loss of direction of individual muscle fibres occur³⁶. Our analysis of both GFP-tagged myosin and electron micrographs in ageing nematodes provides clear evidence of the loosening of sarcomere geometry, the fraying of individual sarcomeres and the loss of direction of individual sarcomeres. *Drosophila* flight muscle also exhibits signs of age-related deterioration³⁷; we therefore speculate that sarcopenia might emerge as another common feature in ageing metazoans.

The *age-1*/PI(3) kinase influences sarcopenia

Disruption of the *age-1*/PI(3) kinase, a downstream signalling component in the DAF-2 insulin-like signalling pathway, delays the onset of GFP-reported nuclear deterioration in muscle and extends healthspan. Our work implicates *age-1*/PI(3) kinase as the first genetic factor influencing nematode sarcopenia and, by analogy, we suggest that PI(3) kinase is a candidate for a similar role in human sarcopenia. Interestingly, *age-1* (*hx546*) effects on the muscle

nuclei are first evident in mid-life, beginning near the post-reproductive transition. Our results reinforce the idea that mid-life events might have a significant impact on healthspan and longevity³⁸. Also noteworthy is the recent finding that *daf-2* (which encodes an insulin-like receptor that influences lifespan²) might normally influence the rate of ageing in several tissues³⁹. Perhaps several components of the insulin-like pathway influence sarcopenia.

How does tissue ageing relate to longevity?

Signals derived from specific tissues that affect the lifespan phenotype have been reported for insulin-like signalling⁴⁰ as well as for the *C. elegans* somatic gonad and germ line^{38,41}. It is not yet clear how these signals influence tissue-specific aspects of ageing, such as muscle decline, nor is it known whether, or how, muscle decline influences longevity. Expression of *age-1* in muscle alone in an otherwise *age-1* mutant background cannot rescue the lifespan extension phenotype, but it can correct lipid metabolism defects⁴⁰. We shall need to determine whether *age-1* expressed solely in muscle rescues sarcopenia and to characterize how the various increased longevity mutants affect tissue-specific aspects of ageing to address the question of how the health of different tissues relates to the lifespan of the organism.

Lack of gene regulation in post-reproductive life

Our analysis of the cell biology of *C. elegans* ageing revealed multiple lines of evidence indicating that gene expression and/or protein turnover is not coordinated with biological need during the post-reproductive period of life. At the ultrastructural level, we confirmed extensive lipid, lipofuscin and yolk accumulation and we noted that the collagenous *C. elegans* cuticle markedly thickens in aged animals. In terms of cell-specific GFP reporters, we noted that all of the 12 reporters we examined remained detectable over the lifespan. Our observations raise the possibility that many genes expressed at the end of the reproductive period might continue to be expressed long into the lifespan. Because selective pressure is thought to decline progressively with age²⁹ and should be effectively non-existent on post-reproductive *C. elegans*, one might predict that biosynthesis might be poorly regulated during post-reproductive life so that a 'left-on' profile of gene expression could endure⁴². Consistent with this, no major changes in *C. elegans* proteins resolved on two-dimensional gels occur over time⁴³. We speculate that non-regulated biosynthesis might contribute to the senescent decline of *C. elegans* by causing the continued production of irrelevant macromolecules. We note that for most LacZ-tagged markers examined in *Drosophila*, genes continue to be expressed into old age^{31,44} (*wg* is an exception⁴⁴). Possibly, the 'expression left-on' phenotype might be a general feature of ageing organisms released from selective pressures.

Even when some aspects of the 'expression left-on' phenotype are operative, increased-longevity mutant *age-1* (*hx546*) exhibits delayed decline in both locomotory activity and nuclear muscle changes. Thus, manipulation of only a subset of age-related phenotypes might suffice to provide significant adjustments to quality of life. Analysis of processes that affect ageing in specific tissues could therefore provide useful clues for combating some deleterious aspects of human ageing. □

Methods

Strains and media

All nematodes were maintained at 20 °C as described⁴⁵. Strains used were: IM175 *unc-119*/GFP (W. Wadsworth, unpublished observations); ZB154 *zdl5* *p_{mec-4}*-GFP (S. Clark, unpublished observations); NW1099 *evl* *b2a* [*P_{unc-129}*/GFP *pM1186*(*dpy-20*(+))]¹³; PD4251 *ccIs4251* (*P_{myo-3}*-GFP, *dpy-20* (+)); RW1596 *myo-3*(*st386*); *P_{myo-3}*-GFP, *rol-6*(*su1006*)¹⁶; DH1033 *bls1* [*YP170*/GFP, *rol-6*(*su1006*)]; *sqt-1*(*sc103*); IAO19 (*P_{col-12}*/GFP, *rol-6*(*su1006*)); EG1653 *oxIs22* (*P_{unc-49}*/GFP; *lin-15*(+)); *rhlIs2* *pat-3*/GFP; TJ1052 *age-1* (*hx546*) II; BE13 *sqt-1*(*sc13*) II; and N2. Double-mutant strains were constructed by standard genetic approaches; *sqt-1* was used as a linked marker to track *age-1* (*hx546*) in crosses; longevity phenotype was confirmed.

Age synchronization and behavioural scoring of ageing nematodes

To age-synchronize nematodes, 5–10 gravid adults were plated on two to four plates to lay eggs for 2–4 h and then removed; the day of egg deposition was scored as day 0. When animals matured to the L4 larval stage, they were distributed (10 nematodes per plate) and scored every 1–3 d for classes A, B and C as described in the text. Animals that did not respond to prodding with a wire and showed no pharyngeal pumping were scored as dead.

Microscopy techniques

After scoring for behavioural class, animals were fixed and embedded by standard methods⁴⁶. Thin sections were collected in transverse or longitudinal aspects and examined on a Philips CM10 transmission electron microscope. Results were drawn from over 500 electron micrographs, drawn primarily from a total of two class A, six class B and three class C animals. Young adult wild-type animals were available from 25 animals in our archival collection. Immunocytochemistry was performed by a post-embedding procedure as described by Paupard *et al.*⁴⁷. Gold-linked secondary antibodies were imaged by electron microscopy on the thin sections of aldehyde-fixed aged animals, using primary antibodies against yolk protein⁴⁷.

For GFP analysis, age-synchronized nematodes were transferred to new plates at 2-day intervals. For each strain, the animals were reared under identical temperature and growth conditions for comparison. For several lines we found that the GFP strains used for these experiments had slightly shorter lifespans than the N2 strain, but we always continued to the end of the lifespan. At several intervals (usually every 2 or 3 days), 5–10 animals were analysed by fluorescence microscopy. Animals were observed and photographed with a higher-magnification fluorescent microscope (Zeiss) with a Real-14 Precision Digital camera attachment. Observations of GFP expression were recorded and colour images were taken for the documentation of results with Magnafire software.

Dye-filling experiments

Stock dye solution (50 µl; 20 mg ml⁻¹ of 5-fluorescein isothiocyanate (FITC) in dimethylformamide) was mixed with 200 µl of M9 buffer and applied to a seeded nematode growth medium plate for 2–12 h before testing⁴⁸. Age-synchronized nematodes were placed on FITC-containing plates for 4–12 h and then transferred to seeded plates without dye for at least 1 h to remove excess FITC from the intestine and from the outside of the animal. Nematodes were mounted and observed as described above for GFP observation, but with a FITC filter.

Received 22 April; accepted 4 September 2002; doi:10.1038/nature01135.

- Vanfleteren, J. R. & Braeckman, B. P. Mechanisms of life span determination in *Caenorhabditis elegans*. *Neurobiol. Aging* **20**, 487–502 (1999).
- Braeckman, B. P., Houthoofd, K. & Vanfleteren, J. R. Insulin-like signaling, metabolism, stress resistance and aging in *Caenorhabditis elegans*. *Mech. Ageing Dev.* **122**, 673–693 (2001).
- Lithgow, G. J. & Walker, G. A. Stress resistance as a determinate of *C. elegans* lifespan. *Mech. Ageing Dev.* **123**, 765–771 (2002).
- Sulston, J. E. & Horvitz, H. R. Post embryonic cell lineages of the nematode *Caenorhabditis elegans*. *Dev. Biol.* **56**, 110–156 (1977).
- White, J. G., Southgate, E., Thomson, J. N. & Brenner, S. The structure of the nervous system of *Caenorhabditis elegans*. *Phil. Trans. R. Soc. Lond. B* **314**, 1–340 (1986).
- Klass, M. R. Aging in the nematode *Caenorhabditis elegans*: major biological and environmental factors influencing life span. *Mech. Ageing Dev.* **6**, 413–429 (1977).
- Bolanowski, M. A., Jacobson, L. A. & Russell, R. L. Quantitative measures of aging in the nematode *Caenorhabditis elegans*. II. Lysosomal hydrolases as markers of senescence. *Mech. Ageing Dev.* **21**, 295–319 (1983).
- Driscoll, M. Cell death in *C. elegans*: Molecular insights into mechanisms conserved between nematodes and mammals. *Brain Pathol.* **6**, 411–425 (1996).
- Johnson, T. E. Aging can be genetically dissected into component processes using long-lived lines of *Caenorhabditis elegans*. *Proc. Natl Acad. Sci. USA* **84**, 3777–3781 (1987).
- Hall, D. H. *et al.* Neuropathology of degenerative cell death in *C. elegans*. *J. Neurosci.* **17**, 1033–1045 (1997).
- Maduro, M. & Pilgrim, D. Identification and cloning of unc-119, a gene expressed in the *Caenorhabditis elegans* nervous system. *Genetics* **141**, 977–988 (1995).
- Mitani, S., Du, H., Hall, D. H., Driscoll, M. & Chalfie, M. Combinatorial control of touch receptor neuron expression in *Caenorhabditis elegans*. *Development* **119**, 773–783 (1993).
- Colavita, A., Krishna, S., Zheng, H., Padgett, R. W. & Culotti, J. G. Pioneer axon guidance by UNC-129, a *C. elegans* TGF-β. *Science* **281**, 706–709 (1998).
- Apfeld, J. & Kenyon, C. Regulation of lifespan by sensory perception in *Caenorhabditis elegans*. *Nature* **402**, 804–809 (1999).
- Perkins, L. A., Hedgecock, E. M., Thomson, J. N. & Culotti, J. G. Mutant sensory cilia in the nematode *Caenorhabditis elegans*. *Dev. Biol.* **117**, 456–487 (1986).
- Campagnola, P. J. *et al.* Three-dimensional high-resolution second-harmonic generation imaging of endogenous structural proteins in biological tissues. *Biophys. J.* **82**, 493–508 (2002).
- Epstein, J., Himmelhoch, S. & Gershon, D. Studies on ageing in nematodes. III. Electron microscopical studies on age-associated cellular damage. *Mech. Ageing Dev.* **1**, 245–255 (1972).
- Hall, D. H. *et al.* Ultrastructural features of the adult hermaphrodite gonad of *Caenorhabditis elegans*: relations between the germ line and soma. *Dev. Biol.* **212**, 101–123 (1999).
- Kimura, K. D., Tissenbaum, H. A., Liu, Y. & Ruvkun, G. *daf-2*, an insulin receptor-like gene that regulates longevity and diapause in *Caenorhabditis elegans*. *Science* **277**, 942–946 (1997).
- Friedman, D. B. & Johnson, T. E. A mutation in the *age-1* gene in *Caenorhabditis elegans* lengthens life and reduces hermaphrodite fertility. *Genetics* **118**, 75–86 (1988).
- Duhon, S. A. & Johnson, T. E. Movement as an index of vitality: comparing wild type and the *age-1* mutant of *Caenorhabditis elegans*. *J. Gerontol. A Biol. Sci. Med. Sci.* **50**, B254–B261 (1995).
- Bolanowski, M. A., Russell, R. L. & Jacobson, L. A. Quantitative measures of aging in the nematode *Caenorhabditis elegans*. I. Population and longitudinal studies of two behavioral parameters. *Mech. Ageing Dev.* **15**, 279–295 (1981).
- Kirkwood, T. B. Evolution of ageing. *Nature* **270**, 301–304 (1977).
- Kirkwood, T. B. Human senescence. *BioEssays* **18**, 1009–1016 (1996).
- Kirkwood, T. B. Changing complexity in aging: a metric not an hypothesis. *Neurobiol. Aging* **23**, 21–22 (2002).
- Johnson, T. E. *et al.* Relationship between increased longevity and stress resistance as assessed through gerontogene mutations in *Caenorhabditis elegans*. *Exp. Gerontol.* **36**, 1609–1617 (2001).
- Finch, C. E. & Kirkwood, T. B. L. *Chance, Development, and Aging* (Oxford Univ. Press, New York, 2000).
- Sozou, P. D. & Kirkwood, T. B. A stochastic model of cell replicative senescence based on telomere shortening, oxidative stress, and somatic mutations in nuclear and mitochondrial DNA. *J. Theor. Biol.* **213**, 573–586 (2001).
- Kirkwood, T. B. Evolution of ageing. *Mech. Ageing Dev.* **123**, 737–745 (2002).
- Leonard, D. S., Bowman, V. D., Ready, D. F. & Pak, W. L. Degeneration of photoreceptors in rhodopsin mutants of *Drosophila*. *J. Neurobiol.* **23**, 605–626 (1992).
- Helfand, S. L. & Napra, B. The expression of a reporter protein, β-galactosidase, is preserved during maturation and aging in some cells of the adult *Drosophila melanogaster*. *Mech. Dev.* **55**, 45–51 (1996).
- Morrison, J. H. & Hof, P. R. Life and death of neurons in the aging brain. *Science* **278**, 412–419 (1997).
- Jucker, M., Bondolfi, L., Calhoun, M. E., Long, J. M. & Ingram, D. K. Structural brain aging in inbred mice: potential for genetic linkage. *Exp. Gerontol.* **35**, 1383–1388 (2000).
- Waterston, R. H. *The Nematode: Caenorhabditis elegans* (ed. Wood, W. B.) 281–337 (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1988).
- Hughes, S. M. & Schiaffino, S. Control of muscle fibre size: a crucial factor in ageing. *Acta Physiol. Scand.* **167**, 307–312 (1999).
- Fujisawa, K. Some observations on the skeletal musculature of aged rats. Part 2. Fine morphology of diseased muscle fibres. *J. Neurol. Sci.* **24**, 447–469 (1975).
- Takahashi, A., Philpott, D. E. & Miquel, J. Electron microscope studies on aging *Drosophila melanogaster*. 3. Flight muscle. *J. Gerontol.* **25**, 222–228 (1970).
- Arantes-Oliveira, N., Apfeld, J., Dillin, A. & Kenyon, C. Regulation of life-span by germ-line stem cells in *Caenorhabditis elegans*. *Science* **295**, 502–505 (2002).
- Garigan, D. *et al.* Genetic analysis of tissue aging in *Caenorhabditis elegans*. A role for heat-shock factor and bacterial proliferation. *Genetics* **161**, 1101–1112 (2002).
- Wolkow, C. A., Kimura, K. D., Lee, M. S. & Ruvkun, G. Regulation of *C. elegans* life-span by insulin like signaling in the nervous system. *Science* **290**, 147–150 (2000).
- Hsin, H. & Kenyon, C. Signals from the reproductive system regulate the lifespan of *C. elegans*. *Nature* **399**, 362–366 (1999).
- Lund, J. *et al.* Global profiles of aging in *Caenorhabditis elegans*. *Curr. Biol.* **12**, 1566–1573 (2002).
- Johnson, T. E. & McCaffrey, G. Programmed aging or error catastrophe? An examination by two-dimensional polyacrylamide gel electrophoresis. *Mech. Ageing Dev.* **30**, 285–297 (1985).
- Rogina, B., Vaupel, J. W., Partridge, L. & Helfand, S. L. Regulation of gene expression is preserved in aging *Drosophila melanogaster*. *Curr. Biol.* **8**, 475–478 (1998).
- Lewis, J. A. & Fleming, J. T. *Caenorhabditis elegans: Modern Biological Analysis of an Organism* (eds Epstein, H. F. & Shakes, D. C.) 4–27 (Academic, San Diego, 1995).
- Hall, D. H. Electron microscopy and three-dimensional image reconstruction. *Methods Cell Biol.* **48**, 395–436 (1995).
- Paupard, M. C., Miller, A., Grant, B., Hirsh, D. & Hall, D. H. Immuno-EM localization of GFP-tagged yolk proteins in *C. elegans* using microwave fixation. *J. Histochem. Cytochem.* **49**, 949–956 (2001).
- Hedgecock, E. M., Culotti, J. G., Thomson, J. N. & Perkins, L. A. Axonal guidance mutants of *Caenorhabditis elegans* identified by filling sensory neurons with fluorescein dyes. *Dev. Biol.* **111**, 158–170 (1985).

Supplementary Information accompanies the paper on Nature's website (<http://www.nature.com/nature>).

Acknowledgements We thank the following people for providing the following strains used in this study: B. Wadsworth for IM175, S. Clark for ZB154, R. Padgett for NW1099, A. Fire for PD4251, P. Hoppe and B. Waterston for RW1596, B. Grant for DH1033, I. Johnstone for IA019, E. Jorgensen for EG1653, and J. Plenefisch for *rhl52*. We also thank the *Caenorhabditis* Genetics Center funded by NIH NCRR for the TJ1052 and BE13 strains. We thank G. Stephey, T. Stephey and K. Nguyen for help with electron microscopy; B. Levine and A. Melendez for helpful discussions about autophagy in *C. elegans*; G. Patterson and B. Grant for critical reading of this manuscript; and T. Johnson and D. Wu for advice about the analysis of behavioural class as a predictor of life expectancy. This work was supported by grants from the National Institute of Neurological Disorders and Stroke and the National Institute on Aging to M.D., and the Center for *C. elegans* Anatomy was supported by a grant from the NIH Division of Research Resources to D.H.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to M.D. (e-mail: driscoll@mbl.rutgers.edu) or D.H.H. (e-mail: hall@accomm.yu.edu).

Experimental realization of the quantum universal NOT gate

F. De Martini*, V. Bužek†‡, F. Sciarrino* & C. Sias*

* Dipartimento di Fisica and Istituto Nazionale di Fisica della Materia, Università "La Sapienza", 00185 Roma, Italy

† Research Center for Quantum Information, Institute of Physics, Slovak Academy of Sciences, Dúbravská cesta 9, 842 28 Bratislava, Slovakia

‡ Department of Mathematical Physics, National University of Ireland, Maynooth, Ireland

In classical computation, a 'bit' of information can be flipped (that is, changed in value from zero to one and vice versa) using a logical NOT gate; but the quantum analogue of this process is much more complicated. A quantum bit (qubit) can exist simultaneously in a superposition of two logical states with complex amplitudes, and it is impossible^{1–3} to find a universal transformation that would flip the original superposed state into a perpendicular state for all values of the amplitudes. But although perfect flipping of a qubit prepared in an arbitrary state (a universal NOT operation) is prohibited by the rules of quantum mechanics, there exists an optimal approximation² to this procedure. Here we report the experimental realization of a universal quantum machine⁴ that performs the best possible approximation to the universal NOT transformation. The system adopted was an optical parametric amplifier of entangled photon states, which also enabled us to investigate universal quantum cloning.

In order to understand the problem of spin flipping, we consider the Poincaré sphere, which represents a state space of a qubit. The points corresponding to $|\Psi\rangle$ and $|\Psi^\perp\rangle$ are antipodes. The desired

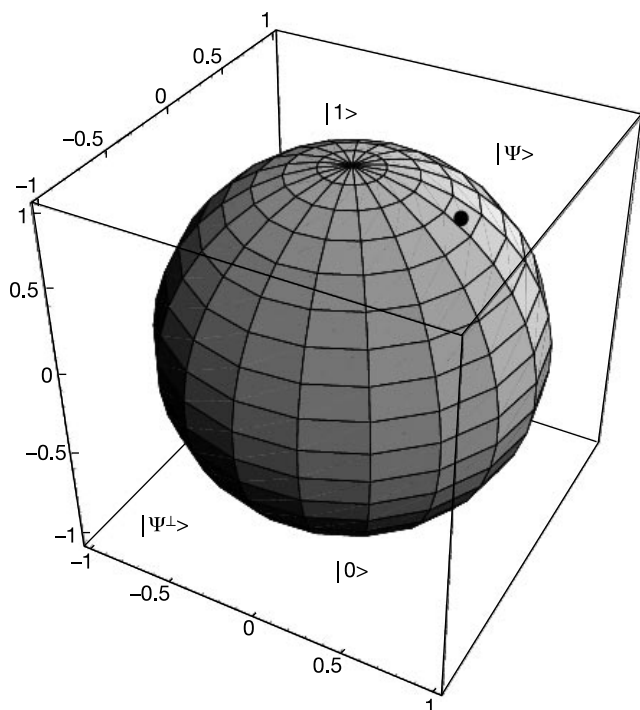


Figure 1 The state space of a qubit is a Poincaré sphere. Pure states are represented by points on the sphere, while statistical mixtures are points inside the sphere. The universal NOT operation corresponds to the inversion of the sphere, because the states $|\Psi\rangle$ and $|\Psi^\perp\rangle$ are antipodes.

spin-flip operation is therefore the inversion of the Poincaré sphere (Fig. 1). It is known that this inversion preserves angles. Therefore, by the arguments of the Wigner theorem the ideal spin-flip operation must be implemented either by a unitary or by an anti-unitary operation. Unitary operations correspond to proper rotations of the Poincaré sphere, whereas anti-unitary operations correspond to orthogonal transformations with determinant equal to -1 . The spin-flip is an anti-unitary operation—that is, it is not completely positive. It is exactly this property that makes the spin-flip operation so important in all criteria of inseparability for two-qubit systems. On the other hand, from above it follows that in the unitary world the ideal universal (U) NOT gate, which would flip a qubit in an arbitrary state, does not exist.

As it is not possible to realize a perfect U-NOT gate which would flip an arbitrary qubit state, it is necessary to investigate what is the best approximation to this gate. This investigation illuminates bounds on information processing imposed by rules of quantum mechanics. There are two possible approaches. The first is based on the measurement of input qubit(s). By using the results of an optimal measurement, one can manufacture an orthogonal qubit, or any desired number of them. In this case, the 'fidelity' (F) of the NOT operation is equal to the fidelity of estimation of the state of the input qubit(s). The second approach would be to approximate an anti-unitary transformation on a Hilbert space of the input qubit(s) by a unitary transformation on a larger Hilbert space which describes the input qubit(s) and ancillas².

The best achievable fidelity of both flipping approaches is the same^{2,3}. That is, the fidelity of the optimal U-NOT gate is equal to the fidelity of the best state-estimation performed on input qubits⁵ (one might say, that in order to flip a qubit we have to transform it into a bit). Even though the fidelity of both processes is the same, the U-NOT gate has a great advantage. Namely, in the measurement-based approach the information encoded in a state of a qubit is irreversibly lost by the measurement process, whereas in the U-NOT gate the information is only redistributed according to specific rules, but owing to the unitarity of the gate it can be recovered.

In our experiment (see below, and Methods) we will consider a flipping of a single qubit. The optimal U-NOT transformation reads²:

$$\hat{U}|\Psi\rangle_a \otimes |X\rangle_{bc} = (2/3)^{1/2} |\Psi\Psi\rangle_{ab} |\Psi^\perp\rangle_c - (1/3)^{1/2} \{|\Psi, \Psi^\perp\rangle\}_{ab} |\Psi\rangle_c \quad (1)$$

where the gate is always prepared in some state $|X\rangle_{bc}$, independently of the input state $|\Psi\rangle$, and $\{|\Psi, \Psi^\perp\rangle\}$ is a symmetric state of two orthogonal qubits. To be specific, equation (1) describes a process

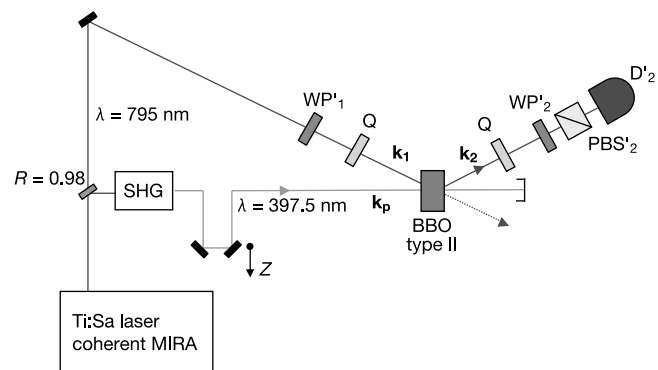


Figure 2 Layout of the apparatus used for the experimental verification of the universality of the optical parametric amplifier (OPA). Various polarization (π)-states of the injected pulse were prepared by a wave-plate WP'_1 (either $\lambda/2$ or $\lambda/4$) inserted on the injection mode k_1 . See Methods for details.

when the original qubit is encoded in system a while the flipped qubit is in system c . The density operator describing the output state of system c is:

$$\sigma^{(\text{out})} = 1/3|\Psi^\perp\rangle\langle\Psi^\perp| + 1/3\mathbf{1} = 2/3|\Psi^\perp\rangle\langle\Psi^\perp| + 1/3|\Psi\rangle\langle\Psi| \quad (2)$$

The identity operator $\mathbf{1}$ in equation (2) reflects the amount of noise induced by the U-NOT gate. The average fidelity of the universal NOT gate is $F = \langle\Psi^\perp|\sigma^{(\text{out})}|\Psi^\perp\rangle = 2/3$. The discussion above describes the case of a single input qubit. When generally N qubits are considered at the input of the U-NOT gate, then the fidelity depends on the number N of input qubits prepared in the state $|\Psi\rangle$. The larger their number, the larger is the fidelity. Specifically, we find that the explicit expression for the fidelity of the U-NOT is $F = (N+1)/(N+2)$, which is exactly the same as the fidelity of the optimal state estimation⁵. We see that in the quantum world governed by unitary operations anti-unitary operations can be performed with a fidelity which is bounded by the amount of classical information available about states of quantum systems.

A natural way to encode a qubit into a physical system is to utilize the polarization states of the photon. In this case, the U-NOT gate can be realized via stimulated emission. The central idea of our experiment is based on the proposal that universal quantum machines⁴ such as the quantum cloner can be realized with the help of stimulated emission in parametric downconversion^{6–8}. Specifically, we consider a qubit to be encoded in a polarization state of a single photon. This photon was injected as the input state into an optical parametric amplifier (OPA) excited by a pulsed, mode-locked ultraviolet (UV) laser beam⁶. The relevant modes of the nonlinear (NL) three-wave interaction driven by the UV pulse were the spatial modes with wavevectors $\mathbf{k}_1$ and $\mathbf{k}_2$ each supporting a horizontal (H) and a vertical (V) linear-polarization (π) of the interacting photons: for example, π_{1H} is the horizontal polarization unit vector associated with $\mathbf{k}_1$. The OPA was frequency degenerate: that is, the interacting photons had the same wavelength $\lambda = 795$ nm. More experimental details are given in Methods. The action of the OPA under suitable conditions, realized in

our case, is described by the ‘squeezing’ hamiltonian⁶: $H_{\text{int}} = i\hbar\chi(\hat{a}_H^\dagger\hat{b}_V^\dagger - \hat{a}_V^\dagger\hat{b}_H^\dagger) + \text{h.c.}$, where χ expresses the nonlinear susceptibility of the active crystal, and h.c. is the hermitian conjugate. Here the field operators $\hat{a}$ and $\hat{b}$ are assumed to be acting on modes $\mathbf{k}_1$ and $\mathbf{k}_2$, respectively. It has been recently shown by theory^{7,8} that H_{int} is invariant under simultaneous general SU(2) transformations of the polarization vectors for modes $\mathbf{k}_1$ and $\mathbf{k}_2$. We may then cast the expression above in the following form:

$$H_{\text{int}} = i\hbar\chi(\hat{a}_\pi^\dagger\hat{b}_{\pi^\perp}^\dagger - \hat{a}_{\pi^\perp}^\dagger\hat{b}_\pi^\dagger) + \text{h.c.} \quad (3)$$

where the field labels refer to the two mutually orthogonal polarization unit vectors for each mode, π and $\pi^\perp$, corresponding respectively to the state vectors: $|\Psi\rangle$ and $|\Psi^\perp\rangle$.

Equation (3) is of central importance in the context of the present work as the amplification efficiency of this type of OPA under any externally injected quantum field (consisting, for example, of a single photon or of a classical ‘coherent’ field) can be made independent of the polarization state of the field. Indeed, it has been shown by a recent experiment on ‘universal quantum cloning’⁹ that the OPA ‘gain’ $g = \chi t$ (where t is time) is independent of any (unknown) polarization state of the injected field: this precisely represents the necessary universality (U) property of the U-NOT gate. We assume that the input photon in the mode $\mathbf{k}_1$ has a polarization π , a condition expressed by the state vector $|\Psi\rangle$, as said. We will describe this polarization state as $\hat{a}_\pi^\dagger|0,0\rangle_{\mathbf{k}_1} = |1,0\rangle_{\mathbf{k}_1}$ where we have used notation introduced in ref. 7: that is, the state $|m,n\rangle_{\mathbf{k}_1}$ represents a state with m photons of the mode $\mathbf{k}_1$ having the polarization π , while n photons have the polarization $\pi^\perp$. Initially there are no excitations in the mode $\mathbf{k}_2$. The initial polarization state of these two modes reads $|1,0\rangle_{\mathbf{k}_1}\otimes|0,0\rangle_{\mathbf{k}_2}$ and it evolves according to the unitary operator $\hat{U} \equiv \exp(-iH_{\text{int}}t/\hbar)$:

$$\hat{U}|1,0\rangle_{\mathbf{k}_1}\otimes|0,0\rangle_{\mathbf{k}_2} = |1,0\rangle_{\mathbf{k}_1}\otimes|0,0\rangle_{\mathbf{k}_2} + g(2^{1/2}|2,0\rangle_{\mathbf{k}_1}\otimes|0,1\rangle_{\mathbf{k}_2} - |1,1\rangle_{\mathbf{k}_1}\otimes|1,0\rangle_{\mathbf{k}_2}) \quad (4)$$

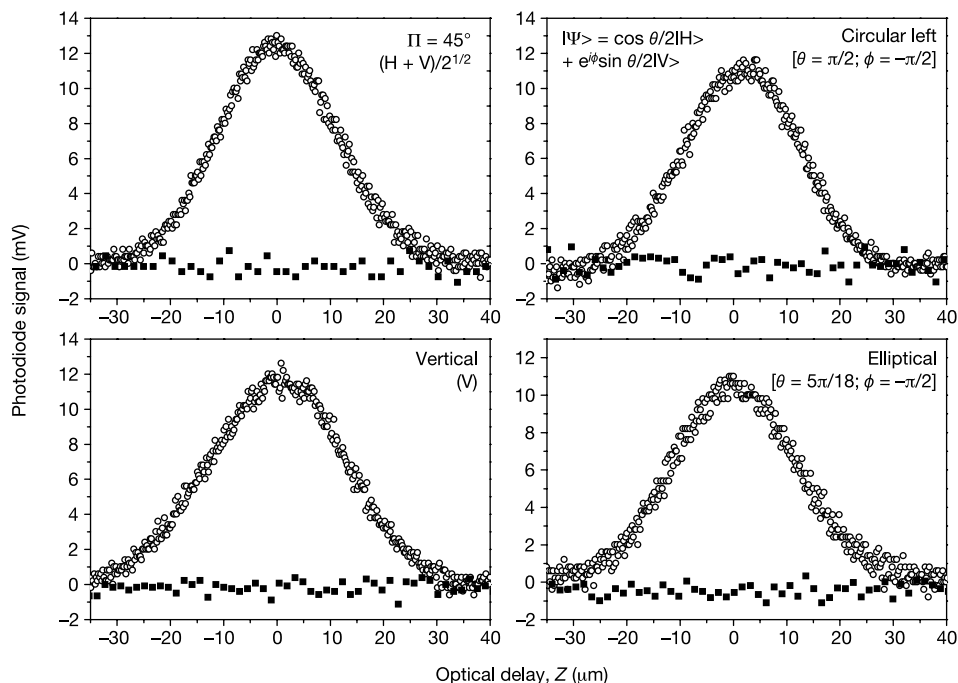


Figure 3 Experimental verification of the universality of the OPA system for different field polarizations, either linear (Π), circular, or generally elliptical. The plots show the amplification pulses detected by D_2 on the OPA output mode, $\mathbf{k}_2$, under the injection on the input mode $\mathbf{k}_1$ of attenuated laser coherent pulses. Each plot corresponds to a state of

definite input polarization π : $\cos(\theta/2)|H\rangle + e^{i\phi}\sin(\theta/2)|V\rangle$ (see text for details). The lack of amplification by detection on the ‘wrong’ polarization channel on mode $\mathbf{k}_2$ is shown by the filled squares.

This approximation for the state vector describing the two modes at times $t > 0$ is sufficient, because the value of g is usually very small in our experiment (see below): $g \ll 1$. The zero-order term corresponds to the process when the input photon in the mode $\mathbf{k}_1$ did not interact in the nonlinear medium, while the second term describes the first-order amplification process. This second term is formally equal (up to a normalization factor) to the right-hand side of equation (2). Here the state $|2,0\rangle_{\mathbf{k}_1}$ describing two photons of the mode $\mathbf{k}_1$ in the polarization state π corresponds to the state $|\Psi\Psi\rangle$. This state vector describes the cloning of the original photon⁷⁻⁹. The vector $|0,1\rangle_{\mathbf{k}_2}$ describes the state of the mode $\mathbf{k}_2$ with a single photon with the polarization orthogonal to π . That is, the output state vector $|\Psi^\perp\rangle = |0,1\rangle_{\mathbf{k}_2}$ represents the flipped version of the input $|\Psi\rangle = |1,0\rangle_{\mathbf{k}_1}$.

To see that the stimulated emission is responsible for creation of the flipped qubit, we compare the state represented by equation (4) with the output state when the vacuum is injected into the NL crystal. In this case, to the same order of approximation as above, we obtain:

$$\hat{U}|0,0\rangle_{\mathbf{k}_1} \otimes |0,0\rangle_{\mathbf{k}_2} = |0,0\rangle_{\mathbf{k}_1} \otimes |0,0\rangle_{\mathbf{k}_2} + g(|1,0\rangle_{\mathbf{k}_1} \otimes |0,1\rangle_{\mathbf{k}_2} - |0,1\rangle_{\mathbf{k}_1} \otimes |1,0\rangle_{\mathbf{k}_2}) \quad (5)$$

We see that the flipped qubit described by the state vector $|0,1\rangle_{\mathbf{k}_2}$ in the right-hand sides of equations (4) and (5) appears with different amplitudes corresponding to the ratio of the probabilities being $R = 2:1$. Our experiment indeed consists of the measurement of R , as we shall see. Note also that the U-NOT operation is not altered by multiplying H_{int} by any overall phase factor.

On the ‘microscopic’ quantum level, the justification of this U property of the OPA amplifier resides in the SU(2) invariance of H_{int} when the spatial orientation of the OPA crystal makes it available for creation by spontaneous parametric downconversion (SPDC) of two-photon entangled ‘singlet’ states^{7,8}. But we note that in the present context the universality property (that is, the π -insensitivity of g) is indeed a ‘macroscopic’ classical feature of the OPA device. So it can be tested equally well by injection of either a quasi-classical radiation state (for example, a coherent field), or of a quantum radiation state (for example, a single-photon Fock state). The test corresponding to injection of an attenuated coherent field is shown in Figs 2 and 3. Details of the apparatus (Fig. 2) are given below.

The universality condition is demonstrated by the plots of Fig. 3 showing the amplification pulses detected by D_2' on the OPA output mode $\mathbf{k}_2$. Each plot corresponds to a definite state of polarization (π) of the field injected on mode $\mathbf{k}_1$: $[\cos(\theta/2)|H\rangle + e^{i\phi}\sin(\theta/2)|V\rangle]$. The polarization was either linear (that is, $\theta = \pi/2$, $\phi = 0$), or circular (that is, $\theta = \pi/2$; $\phi = -\pi/2$), or very generally elliptical: $\theta = 5\pi/18$; $\phi = -\pi/2$. We may check that, in

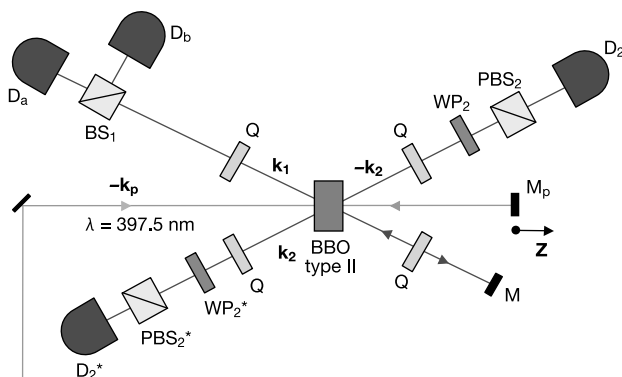


Figure 4 Experimental realization of the U-NOT gate. A single photon, $N = 1$, was injected with a definite π -state into the NL crystal of the OPA along the mode $\mathbf{k}_1$. The π flipping effect was detected on mode $\mathbf{k}_2$. See Methods for details.

spite of the different input π -states, the amplification curves are almost identical. Each coherent pulse injected on the mode $\mathbf{k}_1$ was amplified into an average photon number $M' \approx 5 \times 10^3$ on the output mode $\mathbf{k}_2$.

We now consider the U-NOT gate. By virtue of the tested universality of the OPA amplification, it is sufficient to consider here the OPA injection on mode $\mathbf{k}_1$ by a single photon in just one π -state; for example, in the horizontal π -state: $|\Psi\rangle = |H\rangle$ (Fig. 4).

From the analysis above, it follows that the state of the field emitted by the OPA indeed realizes the U-NOT gate operation: that is, the ‘optimal’ realization of the transformation by flipping of the original qubit originally encoded in the mode $\mathbf{k}_1$. The flipped qubit at the output is in the mode $\mathbf{k}_2$. As shown earlier, the state created by the U-NOT gate cannot be a pure one. There is a minimal amount of noise induced by the process of flipping, which is inevitable in order to preserve complete positiveness of the U-NOT gate. This mixed state is described by the density operator, equation (2). The polarization state of the output photon in the mode $\mathbf{k}_2$ in our experiment is indeed described by this density operator.

The plots of Fig. 5 report our experimental four-coincidence data as function of the time superposition of the UV pump and of the injected single-photon pulses. Our main result consists of the determination of the ratio R between the height of the central peak and that of the flat ‘noise’ contribution. To understand this ratio, we first note that the most efficient stimulation process in the OPA is achieved when a perfect match is achieved: that is, time and space overlap between the UV pump pulse and the optical pulse carrying the input photon. This situation corresponds to the value of the mirror position Z equal to zero. As soon as the mirror is displaced from the position $Z = 0$, the time overlap of the two interacting optical pulses decreases, and the stimulation becomes increasingly less efficient—that is, the spin-flip operation is more noisy. In the limit of large displacements Z , the spin flipping is totally random owing to the fact that the process corresponds to injecting the vacuum field into the crystal. The theoretical ratio between the corresponding probabilities is $R = 2$, as mentioned earlier. In the experiment we have found this ratio to be $R = 1.70 \pm 0.06$. This corresponds, by virtue of equations (4) and (5) above, to a measured value of the fidelity of the U-NOT device: $F = R/(R + 1) = 0.630 \pm 0.008$, to be compared with the theoret-

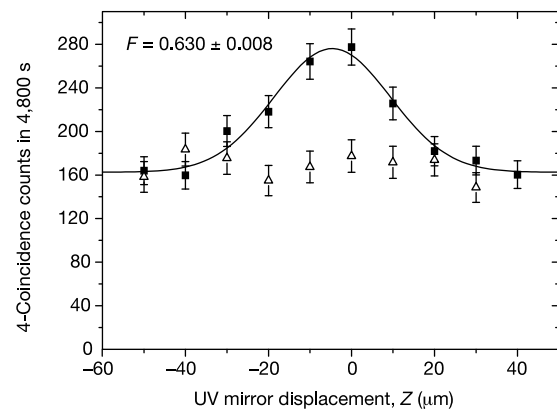


Figure 5 Experimental verification of the optimum conditions of the U-NOT gate. The height of the central peak expresses the rate measured with the π -analyser of mode $\mathbf{k}_2$ set to measure the ‘correct’ vertical (V) polarization, that is, the one orthogonal to the (H) polarization of the input π -state, $|\Psi\rangle = |H\rangle$ of the injected single photon, $N = 1$. By turning by 90° the π -analyser, the amount of the ‘noise’ contribution is represented by a ‘flat’ curve. The noise contribution is provided by the amplification of the unavoidable QED vacuum field on the input mode $\mathbf{k}_1$. Filled squares, plots corresponding to the ‘correct’ polarization; open triangles, plots corresponding to the ‘noise’. The solid line represents the best gaussian fit to the results expressing the ‘correct’ polarization.

cal value of $F = 2/3 = 0.666$. Note that the height of the central peak in Fig. 5 does not decrease towards zero for large values of Z , as expected. This effect, due to instrumental imperfections, is discussed in Methods.

By a different measurement configuration, the apparatus was adopted to investigate the process of 'quantum cloning' of an $N = 1$ input qubit into $M = 2$ output qubits on mode k_1 . We have confirmed the results of ref. 9 by attaining a cloning fidelity $F = 0.810 \pm 0.008$ (theoretical value, $F = 5/6 \approx 0.833$). □

Methods

Basic optical equipment

The main source of all experiments was a Ti:Sa coherent MIRA mode-locked pulsed laser, providing (by second harmonic generation, SHG) the 'pump' field for the quantum-injected OPA associated with the spatial mode having wavevector k_p and wavelength $\lambda_p = 397.5$ nm (that is, in the UV range of the spectrum⁶). The time duration of each UV pulse was $\tau \approx 140$ femtoseconds (fs). The OPA active element, consisting of a 1.5-mm-thick NL crystal of β -barium borate (BBO) cut for type II phase-matching, was able to generate by SPDC—that is, by amplification of the vacuum field—linear polarization (π)-entangled pairs of photons. The 'intrinsic phase' of the OPA was set so as to generate by SPDC 'singlet' entangled states on the output optical modes, k_1, k_2 (ref. 10). The photons of each pair were emitted with equal wavelengths $\lambda = 795$ nm over two spatial modes k_1 and k_2 making an external angle of 8° . In all experiments, the time (t) optical walk-off effects due to the birefringence of the NL crystal were compensated by inserting in the modes k_1 and k_2 fixed X-cut, 4.8-mm-thick quartz plates (Figs 2 and 4). All adopted photodetectors (D), but D_2 , were equal SPCM-AQR14 Si-avalanche nonlinear single-photon units with nearly equal quantum efficiencies ~ 0.55 . One interference filter with bandwidth $\Delta\lambda = 6$ nm was placed in front of each detector D. Only the detector D_2 (Fig. 2) was a linear photodiode. Polarizing beam-splitters (PBS) in Figs 2 and 4 are adopted from measurement devices providing polarization analysis.

Universality test

See Figs 2 and 3. Because the universality (U) of the OPA transformation (that is, the insensitivity of g to the (unknown) π state of any input photon) is a macroscopic 'classical' property of the OPA device, the U-test has been carried out by injection of the strongly attenuated laser beam, with wavelength $\lambda = 795$ nm, contributed via a beam splitter by the main mode-locked source and directed along the OPA injection mode k_1 . The parametric amplification, with $g \approx 0.11$, was detected at the OPA output mode k_2 by the linear Si photodiode SGD100 (D_2), filtered by an interference filter with bandwidth $\Delta\lambda = 3$ nm. The time superposition in the NL crystal of the 'pump' and of the 'injection' pulses was assured by micrometric displacements (Z) of a two-mirror optical 'trombone'. The pulse shapes shown by the coincidence data (Figs 3 and 5) as function of Z are indeed the signature for actual amplification: that is, they arise from the effective time and space superposition in the NL crystal of the UV 'pump' pulse and of the optical pulses with $\lambda = 795$ nm, injected into the OPA. Relevant, different π -states of the injected field, formally expressed by the captions of the panels of Fig. 3, were prepared by a single wave-plate WP_1 , corresponding to a suitable optical retardation, equal to $\lambda/2$ or to $\lambda/4$, between the two orthogonal basis π -states, that is, horizontal (H) and vertical (V). The OPA amplified output π -states were detected by an apparatus inserted on mode k_2 and consisting of the set ($WP_2 + \pi$ -analyser), the last device being provided by the polarizing beam splitter PBS_2 .

Realization of the U-NOT gate

See Figs 4 and 5. The UV pump beam, back-reflected by a spherical mirror M_p with 100% reflectivity and micrometrically adjustable-position Z , excited the NL OPA crystal amplifier in both directions ($-k_p$) and k_p , correspondingly oriented towards the right (R) and the left (L) sides of Fig. 4. An SPDC process excited by the $-k_p$ pump mode created single photon-pairs with wavelength $\lambda = 795$ nm in entangled singlet π -states. One photon of each pair, emitted over $-k_1$, was reflected by a spherical mirror M onto the NL crystal where it provided the $N = 1$ quantum injection into the OPA excited by the UV pump beam associated with the back-reflected mode k_p . Consider the qubit flipping related to the OPA injection of a single photon ($N = 1$) over the mode k_1 in the state $|\Psi\rangle = |H\rangle$. Because of the low pump intensity, the $N = 2$ photon injection probability has been evaluated to be $\sim 10^{-2}$ smaller than the one for $N = 1$. The twin photon emitted over $-k_2$, π -selected by the devices ($WP_2 + PBS_2$) and detected by D_2 , provided the 'trigger' of the overall experiment. Because of the EPR non-locality implied by the singlet state, the π -selection on mode $-k_2$ provided the realization on k_1 of the state $|\Psi\rangle = |H\rangle$ of the injected photon. The field's π -state on the output mode k_2 was analysed by the optical device combination ($WP_2 + PBS_2$) and measured by D_2 . Detectors D_a, D_b were coupled to the field associated with the mode k_1 via a normal beam splitter, BS. The rate of the four-coincidences involving all detectors (D_2^*, D_2, D_a, D_b) was experimentally measured by an electronic four-coincidence apparatus having a time resolution of 3 ns. Note that the height of the central peak in Fig. 5 does not decrease towards zero for large values of $|Z|$, as expected. This effect is attributable entirely to the limited time resolution of the four-coincidence apparatus. The effect would disappear if the resolution could be pushed into the sub-picosecond range, precisely of the order of the time duration of the interacting pump and injection pulses: $\tau' \approx 140$ fs. Such a resolution is hardly obtainable with the present technology.

Received 31 May; accepted 23 August 2002; doi:10.1038/nature01093.

1. Bechmann-Pasquinucci, H. & Gisin, N. Incoherent and coherent eavesdropping in the six-state protocol of quantum cryptography. *Phys. Rev. A* **59**, 4238–4248 (1999).
2. Bužek, V., Hillery, M. & Werner, R. F. Optimal manipulations with qubits: Universal-NOT gate. *Phys. Rev. A* **60**, R2626–R2629 (1999).
3. Gisin, N. & Popescu, S. Spin flips and quantum information for antiparallel spins. *Phys. Rev. Lett.* **83**, 432–435 (1999).
4. Alber, G. et al. *Quantum Information: An Introduction to Basic Theoretical Concepts* (Springer Tracts in Modern Physics Vol. 173, Springer, Berlin, 2001).
5. Derka, R., Bužek, V. & Ekert, A. K. Universal algorithm for optimal estimation from finite ensembles via realizable generalized measurement. *Phys. Rev. Lett.* **80**, 1571–1575 (1998).
6. De Martini, F. Amplification of quantum entanglement. *Phys. Rev. Lett.* **81**, 2842–2845 (1998).
7. Simon, C., Weihs, G. & Zeilinger, A. Optimal quantum cloning by stimulated emission. *Phys. Rev. Lett.* **84**, 2993–2996 (2000).
8. De Martini, F., Mussi, V. & Bovino, F. Schroedinger cat states and optimal universal quantum cloning by parametric amplification. *Opt. Commun.* **179**, 581–589 (2000).
9. Lamas-Linares, A., Simon, C., Howell, J. C. & Bouwmeester, D. Experimental quantum cloning of single photons. *Science* **296**, 712–714 (2002).
10. De Martini, F., Di Giuseppe, G. & Padua, S. Multiparticle quantum superpositions and stimulated entanglement by parity selective amplification of entangled states. *Phys. Rev. Lett.* **87**, 150401 (2001).

Acknowledgements We thank M. Hillery, C. Simon, S. Popescu, M. D'Ariano for discussions, and V. Mussi, A. Mazzei, F. Bovino, S. Padua for early experimental collaboration. This work was supported by the FET European Networks ATEST, EQUIP, QUBITS, MURST, and INFN.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to F.D.M. (e-mail: francesco.demartini@uniroma1.it).

Observation of coupled magnetic and electric domains

M. Fiebig*†, Th. Lottermoser*, D. Fröhlich*, A. V. Goltsev‡ & R. V. Pisarev‡

* Institut für Physik, Universität Dortmund, 44221 Dortmund, Germany

† Max-Born-Institut, Max-Born-Straße 2A, 12489 Berlin, Germany

‡ Ioffe Physical Technical Institute of the Russian Academy of Sciences, 194021 St Petersburg, Russia

Ferroelectromagnets are an interesting group of compounds that complement purely (anti-)ferroelectric or (anti-)ferromagnetic materials—they display simultaneous electric and magnetic order^{1–3}. With this coexistence they supplement materials in which magnetization can be induced by an electric field and electrical polarization by a magnetic field, a property which is termed the magnetoelectric effect⁴. Aside from its fundamental importance, the mutual control of electric and magnetic properties is of significant interest for applications in magnetic storage media and 'spintronics'^{2,3}. The coupled electric and magnetic ordering in ferroelectromagnets is accompanied by the formation of domains and domain walls. However, such a cross-correlation between magnetic and electric domains has so far not been observed. Here we report spatial maps of coupled antiferromagnetic and ferroelectric domains in $Y\text{MnO}_3$, obtained by imaging with optical second harmonic generation. The coupling originates from an interaction between magnetic and electric domain walls, which leads to a configuration that is dominated by the ferroelectromagnetic product of the order parameters.

The ferroelectromagnetic (FEM) manganites RMnO_3 with $R \in \{\text{Sc}, \text{Y}, \text{In}, \text{Ho-Lu}\}$ are multiple-order-parameter compounds with four 180° domains denoted by $(+P, +I)$, $(+P, -I)$, $(-P, +I)$, $(-P, -I)$ ⁵. Here, $\pm P$ is the independent component of the ferroelectric (FEL) order parameter, a polar vector, which is invariant under time reversal⁶ and describes the breaking of the inversion symmetry due to the FEL polarization along the hexagonal z axis below the Curie temperature $T_C = 570\text{--}990$ K⁷. On the other hand,

$\pm l$ is the independent component of the antiferromagnetic (AFM) order parameter, a third-rank axial tensor⁶, which is not invariant under time reversal⁶, thus describing the breaking of time-reversal symmetry by the triangular AFM ordering of the Mn spins in the basal xy plane below the Néel temperature $T_N = 70\text{--}130\text{ K}$. The crystallographic and magnetic unit cells of the compounds are presented in refs 9–11.

With linear optics the observation of 180° domains is not possible. However, it will be seen that with nonlinear optics both electric and magnetic 180° domains can be distinguished in a single experimental set-up. The simplest nonlinear optical process is second harmonic generation (SHG) in which light with the frequency ω is incident on a crystal, inducing an electromagnetic polarization at the frequency 2ω , which acts as source for a second harmonic light wave emitted from the crystal. The coupling of the source term $S(2\omega)$ to the incident light wave $E(\omega)$ by a susceptibility $\hat{\chi}$ can be expressed by an expansion describing its dependence on the FEL and AFM order parameters¹². In the (bi)linear approximation we obtain

$$S(2\omega) = \epsilon_0(\hat{\chi}(0) + \hat{\chi}(P) + \hat{\chi}(l) + \hat{\chi}(Pl))E(\omega)E(\omega). \quad (1)$$

Following the Neumann principle, the set of non-zero tensor components of $\hat{\chi}$ is determined by the symmetry of the paraelectric paramagnetic lattice for $\hat{\chi}(0)$, the FEL charge lattice for $\hat{\chi}(P)$, the AFM spin lattice for $\hat{\chi}(l)$, and the ‘complete’ FEM lattice for $\hat{\chi}(Pl)$. Above T_c , only $\hat{\chi}(0)$ is allowed because $P = l = 0$. However, owing to the inversion symmetry of the crystal above T_c even $\hat{\chi}(0)$ is zero in the leading order so that SHG is not observed. In the range $T_c > T > T_N$, FEL contributions to SHG from $\hat{\chi}(P)$ are allowed⁶. Below T_N , the FEL contributions should remain unchanged while being supplemented by AFM contributions from $\hat{\chi}(l)$ and by FEM contributions from $\hat{\chi}(Pl)$. With the (bi)linear dependence on the order parameters in equation (1) we obtain $\hat{\chi}(-O) = -\hat{\chi}(O)$ (with $O \in \{P, l, Pl\}$). For opposite magnetic electric domains this leads to an opposite sign of $S(2\omega)$ corresponding to a 180° phase difference of the respective second harmonic waves. By interference with a reference wave at the frequency 2ω the phase difference can be converted into an intensity difference which allows the distinguishing of 180° domains by SHG in contrast to linear optics^{13,14}.

The most intensely investigated and technically promising¹⁵

compound from the FEM RMnO₃ series is YMnO₃. We investigated various polished YMnO₃ platelets (thickness, $\sim 100\text{ }\mu\text{m}$) oriented perpendicular to one of the cartesian axes x, y, z . Both the intensity and the phase of the second harmonic light were measured with a transmission set-up, using the previously described interference technique^{13,14}. For all combinations of i polarized incoming light and j polarized second harmonic light ($i, j \in \{x, y, z\}$) the spectral dependence of the SHG was determined; this leads to the four independent spectra shown in Fig. 1. According to Table 1 the three spectra shown as grey lines correspond to different FEL tensor components of $\hat{\chi}(P)$ in equation (1). They are already present at room temperature and only observed for a wavevector $\mathbf{k} \parallel \mathbf{z}$ of the incident light. The fourth spectrum in Fig. 1b appears only below T_N and is thus induced by the magnetic ordering. According to Table 1 only FEM contributions from $\hat{\chi}(Pl)$ can explain the presence of such a signal for $\mathbf{k} \parallel \mathbf{z}$. Table 1 further predicts that for this case the same contribution must be observable for $\mathbf{k} \parallel \mathbf{x}$; this is clearly confirmed by Fig. 1a. Purely AFM contributions are present for $\mathbf{k} \parallel \mathbf{z}$ but absent for $\mathbf{k} \parallel \mathbf{x}$. Such contributions are not detected, thus excluding a second harmonic process of the $\hat{\chi}(l)$ type. Therefore the second harmonic spectrum in Fig. 1a manifests the first example of a nonlinear optical process coupling simultaneously to an electric and a magnetic order parameter, which is termed FEM-SHG in the following.

In Fig. 2 the spatially resolved second harmonic light reveals the different domain structures of the sample in three independent experiments. Figure 2a shows the FEL domain structure gained with the second harmonic signal from $\hat{\chi}(P)$. Domains with $\pm P$ appear as bright and dark areas, respectively, owing to the aforementioned interference technique^{13,14}. In the same way Fig. 2b was obtained with the second harmonic light from $\hat{\chi}(Pl)$. With the bilinear coupling to P and l , any FEL domain wall in Fig. 2a should also be present in Fig. 2b because of the related change of sign of the product Pl and the resulting 180° phase shift of the second harmonic wave. However, the domain structure in Fig. 2b does not show any of the FEL walls, but appears to be independent of the domain structure in Fig. 2a. This is only possible if any reversal of the FEL order parameter P is accompanied by a simultaneous change of the AFM order parameter l , so that the FEM product Pl retains its sign. This coupling of order parameters is confirmed by Fig. 2c, which shows the domain structure obtained from the interference of the second harmonic contributions from $\hat{\chi}(P)$ and $\hat{\chi}(Pl)$. In this image bright and dark regions correspond to different AFM domains with $\pm l$. (A change of sign of P affects both second harmonic contributions and does not change their relative phase.) As we suspected, the AFM order parameter in Fig. 2c changes its orientation whenever a FEL domain wall is crossed. There are no solely FEL walls (which is why there are no red lines) in Fig. 2d. Only solely AFM walls, that is, reversals of the AFM order parameter inside a FEL domain, are observed. These are the magnetic walls observed with FEM-SHG in Fig. 2b. Only these walls change upon every temperature cycle through T_N , whereas the AFM walls coupled to FEL walls are reproducible until the FEL domains themselves are changed by poling at room temperature. The peculiar coupling of the magnetic to the electric order parameter thus means that the independent magnetic domain configuration is determined by the FEM product

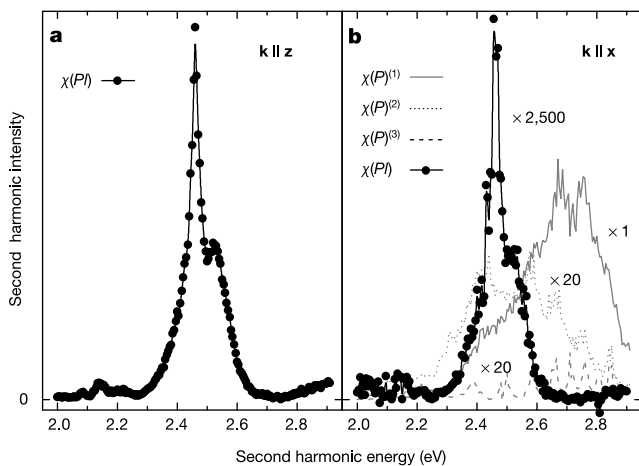


Figure 1 Second harmonic spectra of FEM YMnO₃ at 6 K. **a**, With light incident along the hexagonal z axis, only a magnetic second harmonic signal is observed. **b**, With light incident along the x axis, three independent components $\hat{\chi}(P)^{(1,2,3)}$ from the FEL ordering are observed⁶. The presence of a magnetic second harmonic signal for $\mathbf{k} \parallel \mathbf{x}$ with the same spectral dependence as for $\mathbf{k} \parallel \mathbf{z}$ reveals that only FEM SHG from $\hat{\chi}(Pl)$ can be the origin of the additional second harmonic signal below T_N while $\hat{\chi}(l)$ contributions coupling exclusively to the AFM order are not present.

Table 1 Source term for SHG in ferroelectromagnetic YMnO₃

Source term	$S(P)$	$S(l)$	$S(Pl)$
$\mathbf{k} \parallel \mathbf{z}$	0	S_l	S_{Pl}
$\mathbf{k} \parallel \mathbf{x}$	S_P	0	S_{Pl}

The source term S for the leading second harmonic contributions coupling to P , l , and Pl is derived on the basis of equation (1) and the symmetries of the FEL, the AFM and the FEM lattices, respectively (see text). For this purpose, selection rules for $\hat{\chi}$ from ref. 6 have to be used. The independent components are denoted as $S_{P,l,Pl}$. SHG, second harmonic generation; FEL, ferroelectric; AFM, antiferromagnetic; FEM, ferroelectromagnetic.

Pl rather than by the AFM order parameter l . The domain pattern in Fig. 2b represents a configuration in which two FEM pseudo-domains are distinguished by an opposite sign of the product Pl with no distinction being made between the domain states with $(+P, +l)$ and $(-P, -l)$, or $(+P, -l)$ and $(-P, +l)$.

Calculations have already showed¹⁶ that electronic excitations of the Mn^{3+} ions can lead to bulk second harmonic contributions coupling simultaneously to P and l . However, in order to explain the coupling of domains, a direct coupling between the FEL and AFM order parameters is required. The linear magnetoelectric effect, which represents the simplest manifestation of such a bulk coupling, is symmetry-forbidden in $YMnO_3$ (refs 4, 6). Thus, a coupling between only the electric and magnetic domain walls becomes a leading effect determining the bulk distribution of domains¹⁷.

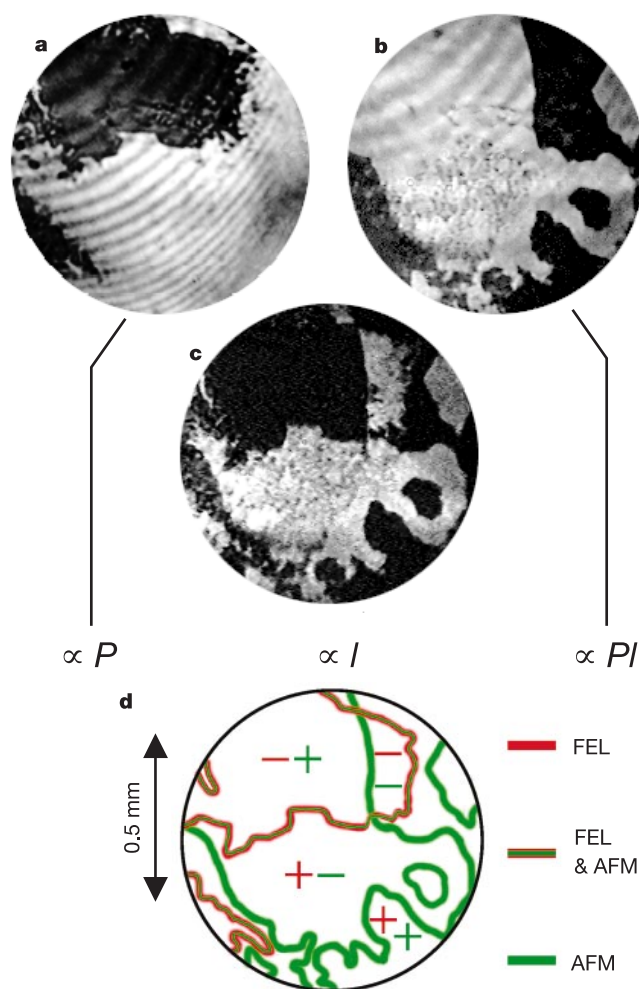


Figure 2 Coexisting electric and magnetic domains of an $YMnO_3$ sample at 6 K imaged with second harmonic light at 2.46 eV. **a**, Exposed with second harmonic light from $\chi(P)$. Dark and bright areas correspond to opposite FEL domains. **b**, Exposed with FEM-second harmonic light from $\chi(Pl)$. Bright and dark regions are distinguished by an opposite sign of the product Pl so that states with $(+P, +l)$ and $(-P, -l)$, or $(+P, -l)$ and $(-P, +l)$, are indistinguishable. The resulting two domain configurations may therefore be termed FEM (pseudo-) domains. **c**, Exposed with interfering second harmonic contributions from $\chi(P)$ and $\chi(Pl)$. Dark and bright areas correspond to opposite AFM domains. **d**, Topology of FEL (red) walls and AFM (green) walls in the sample with $\pm$ signs of the corresponding colour indicating the orientation of the FEL and AFM order parameters in selected domains. Solely red walls are absent because any FEL wall is accompanied by an AFM wall coupled to it. This is why the sign of the FEM product Pl is conserved upon crossing a FEL domain border, and FEL walls are invisible in **b**.

Macroscopically the opposite polarization in neighbouring FEL domains induces a strain σ_{jk} at the FEL wall. On the other hand the gradual reorientation of the Mn spins at the AFM wall induces a local magnetization M_i . For overlapping FEL and AFM walls the piezomagnetic effect thus leads to a non-zero contribution $F = q_{ijk}M_i\sigma_{jk}$ to the free energy. Here, q_{ijk} denotes the components of the piezomagnetic tensor that are allowed in $YMnO_3$ and parametrize the (yet unspecified) microscopic mechanisms contributing to the coupling of domains^{17,18}. The observed coupling can therefore be understood as a consequence of the pinning or creation of AFM domain walls at the position of the FEL walls owing to the reduction of the total energy of the FEM crystal.

Thus we have observed FEM-SHG and a domain configuration dominated by the FEM product Pl of the FEL and AFM order parameters P and l in bulk $YMnO_3$. The underlying very strict coupling of order parameters is important from the point of view of fundamental physics. But a coupling of magnetic and electric order parameters also has possibilities for an application in spintronic devices. The ability to couple to the electric and/or the magnetic long-range order opens up an additional degree of freedom in device construction. This may for instance be used in the design of new data-storage media as an alternative to magneto-optical disks by replacing the slow magnetic writing process by a fast magnetization reversal through electric poling. Furthermore, in spin valves composed from one ferromagnetic and one FEM layer, one of the layers would no longer need to be pinned if electric writing and magnetic reading were used. FEM compounds with AFM spin ordering may permit us to control the degree of exchange-biasing by controlling the topology of the AFM order parameter with an electric field. These developments will be supported by the invention of new FEM compounds with high transition temperatures¹. □

Received 23 July; accepted 16 August 2002; doi:10.1038/nature01077.

- Smolenskii, G. A. & Chupis, I. E. Ferroelectromagnets. *Sov. Phys. Usp.* **25**, 475–493 (1982).
- Schmid, H. *Ferroelectrics* **162**, 317–338 (1994).
- Hill, N. A. Why are there so few magnetic ferroelectrics? *J. Phys. Chem. B* **104**, 6694–6709 (2000).
- O'Dell, T. H. *The Electrostatics of Magneto-Electric Media* (North-Holland, Amsterdam, 1970).
- Aizu, K. Possible species of ferromagnetic, ferroelectric, and ferroelastic crystals. *Phys. Rev. B* **2**, 754–772 (1970).
- Birss, R. R. *Symmetry and Magnetism* (North-Holland, Amsterdam, 1966).
- Yakel, H. L., Koehler, W. C., Bertaut, E. F. & Forrat, E. F. On the crystal structures of the manganite (III) trioxides of the heavy lanthanides and yttrium. *Acta Crystallogr.* **16**, 957–962 (1963).
- Bertaut, E. F. & Mercier, M. Structure magnetique de $MnYO_3$. *Phys. Lett.* **5**, 27–29 (1963).
- van Aken, B. B., Meetsma, A. & Palstra, T. T. M. Hexagonal $YMnO_3$. *Acta Crystallogr. C* **57**, 230–232 (2001).
- van Aken, B. B. *Structural Response to Electronic Transitions in Hexagonal and Ortho-Manganites* PhD thesis, Rijksuniversiteit Groningen (2001).
- Fiebig, M. et al. Determination of the magnetic symmetry of hexagonal manganites by second harmonic generation. *Phys. Rev. Lett.* **84**, 5620–5623 (2000).
- Sa, D., Valenti, R. & Gros, C. A generalized Ginzburg-Landau approach to second harmonic generation. *Eur. Phys. J. B* **14**, 301–305 (2000).
- Fiebig, M., Fröhlich, D., Leute, S. & Pisarev, R. V. Topography of antiferromagnetic domains using second harmonic generation with an external reference. *Appl. Phys. B* **66**, 265–270 (1998).
- Leute, S., Lottermoser, Th. & Fröhlich, D. Nonlinear spatially resolved phase spectroscopy. *Opt. Lett.* **24**, 1520–1522 (1999).
- Fujimura, N., Ishida, T., Yoshimura, T. & Ito, T. Epitaxially grown $YMnO_3$ film: new candidate for nonvolatile memory devices. *Appl. Phys. Lett.* **69**, 1011–1013 (1996).
- Iizuka-Sakano, T., Hanamura, E. & Tanabe, Y. Second-harmonic-generation spectra of the hexagonal manganites $RMnO_3$. *J. Phys. Cond. Matter* **13**, 3031–3055 (2001).
- Soboleva, T. K. & Stefanovskii, E. P. Equilibrium state of the magnetic subsystem of antiferromagnetic ferroelectric. *Sov. Phys. Solid State* **25**, 1637–1638 (1983).
- Filippetti, A. & Hill, N. A. First-principles study of structural, electronic and magnetic interplay in ferroelectromagnetic yttrium manganite. *J. Magn. Magn. Mater.* **236**, 176–1898 (2001).

Acknowledgements We thank Y. Tanabe, E. Hanamura and K. Hagita for discussions, and the Deutsche Forschungsgemeinschaft and the Alexander-von-Humboldt-Stiftung for financial support.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to M.F. (e-mail: fiebig@mbi-berlin.de).

Millennial-scale storminess variability in the northeastern United States during the Holocene epoch

Anders J. Noren*, Paul R. Bierman*, Eric J. Steig†, Andrea Lini* & John Southon‡§

* Department of Geology, University of Vermont, Burlington, Vermont 05405, USA

† Quaternary Research Center and Department of Earth and Space Sciences, University of Washington, Seattle, Washington 98195, USA

‡ Center for Accelerator Mass Spectrometry, Lawrence Livermore National Laboratory, Livermore, California 94551, USA

For the purpose of detecting the effects of human activities on climate change, it is important to document natural change in past climate¹. In this context, it has proved particularly difficult to study the variability in the occurrence of extreme climate events, such as storms with exceptional rainfall¹. Previous investigations have established storm chronologies using sediment cores from single lakes^{2–8}, but such studies can be susceptible to local environmental bias. Here we date terrigenous inwash layers in cores from 13 lakes, which show that the frequency of storm-related floods in the northeastern United States has varied in regular cycles during the past 13,000 years (13 kyr), with a characteristic period of about 3 kyr. Our data show four peaks in storminess during the past 14 kyr, approximately 2.6, 5.8, 9.1 and 11.9 kyr ago. This pattern is consistent with long-term changes in the average sign of the Arctic Oscillation⁹, suggesting that modulation of this dominant atmospheric mode may account for a significant fraction of Holocene climate variability in North America and Europe.

Lakes in the hilly terrain of Vermont and eastern New York contain sedimentary archives that consist of organic lake mud (gyttja), punctuated by layers of terrestrially derived material². These terrigenous layers are commonly coarser, less organic, and contain more macrofossils of terrestrial plants than the surrounding gyttja; they are graded² and have distinctive stable isotopic signatures². Previous work has demonstrated that such terrigenous layers are deposited by exceptional runoff events when rainfall of great intensity and/or duration affects mountainous lake drainage basins^{2–8}. In New England, the heaviest rains occur during localized convective storms, “nor’easters” or other mid-latitude cyclones, and (more rarely) tropical storms, hurricanes, or their remnants. During and immediately after these events, material stored in upland streams and on steep basin hillslopes is eroded and transported to lake basins—as indicated by our field observations during two major storms, during which we documented increased fluvial transport of woody forest debris, sand and gravel into lakes. Thus, stratigraphic analysis and dating of lake sediment cores allows the determination of palaeostorm chronologies^{2–8}.

Other mechanisms (earthquakes, fires, lake-level fluctuations, and removal of vegetation by drought or disease) may also cause or facilitate the deposition of terrigenous material in lakes, but the effects of such other forcings have been shown to be minimal in New England^{2,3}. The first European settlers in the area deforested most of the landscape, but the significant effects of these and other human activities are limited to the past ~250 yr (ref. 3). Snowmelt floods do not transport enough sediment to cause the deposition of terrigenous layers^{5,8}. Analyses of cores from lakes with small, low-relief drainage basins show that periods of low aquatic primary productivity have not occurred in this region during the Holocene³.

Analyses of nine alluvial fans near our coring sites show periods of aggradation caused by increased storminess, runoff and hillslope erosion^{3,10} correlated in time with the lacustrine storm records we present here. Most relevant to the argument that terrigenous layers in lake sediment represent storms are findings in locations with long documentary records, which reveal a strong correlation between heavy rainfall and the occurrence and thickness of terrigenous layers in lacustrine sediment^{4–7}.

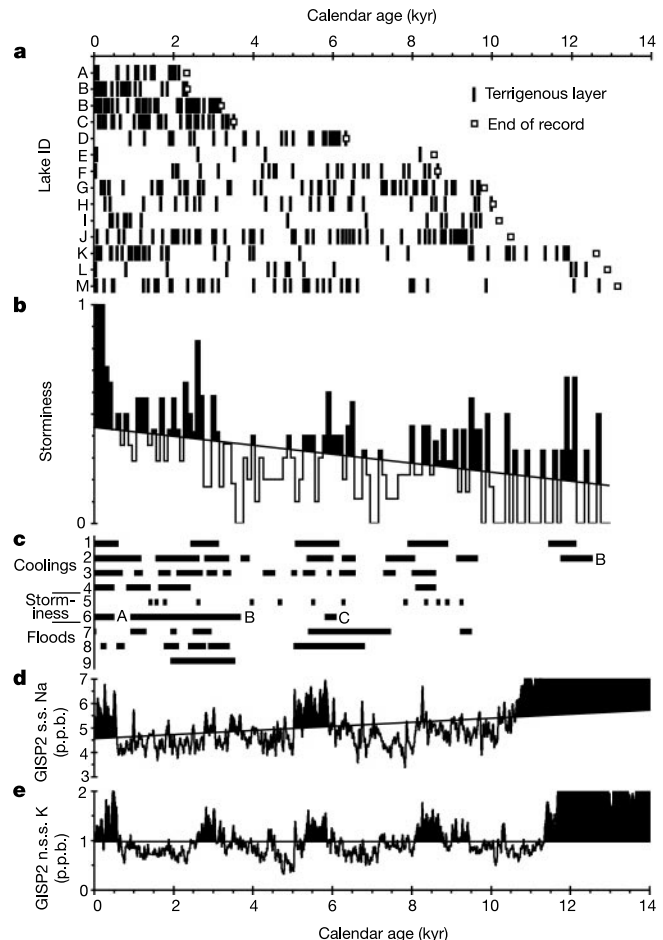


Figure 1 Inferred storminess in the northeastern United States and relevant climate records. **a**, Individual terrigenous sedimentation event chronologies from study lakes. Letters correspond to lakes in Supplementary Information map and table. Replicate cores recovered in each of four basins contained similar stratigraphies. Two cores were recovered from one lake, each in a distinct depocentre. These depocentres record terrigenous sediment delivery from different stream basins. **b**, Histogram of terrigenous sedimentation events (100-yr bins). Values above the superimposed linear regression are shaded. Histogram values are weighted by the inverse of the number of chronologies that cover each time interval. The trend in the data may reflect the slow progradation (growth) of the stream deltas into the lakes, toward the coring locations. As these deltas approach, smaller pulses of sediment are more readily transported to the coring locations. **c**, Other relevant climate records. Coolings as follows. 1, GISP2 glaciochemical cold events²¹. 2, Glacial expansions in the Alps²³; 2B, the Younger-Dryas event¹⁵. 3, Glacial expansions in Scandinavia^{24,25}. 4, Norwegian glacial expansions²⁶. 5, Storm-related aggradational events on alluvial fans in Vermont^{3,10}. Storminess as follows. 6A, LIA historical evidence¹⁸; 6B, high frequency of hurricane landfalls along the northern Gulf of Mexico coast¹⁹; 6C, occurrence of storm surge deposits on the northwestern coast of England²⁰. Floods as follows. 7, Increased magnitude of 1.58-yr recurrence interval floods in the north-central United States (NCUS)¹⁷. 8, Increased magnitude of the largest floods in the NCUS¹⁷. 9, Highest frequency of megafloods on the Mississippi River¹⁶. **d**, GISP2 sea-salt (s.s.) Na and **e**, GISP2 non-sea-salt (n.s.s.) K concentrations, with values above the superimposed linear regressions shaded²¹. Other GISP2 aerosol deposition time series²¹ exhibit variability similar to s.s. Na and n.s.s. K.

§ Present address: Earth System Science, University of California, Irvine, California 92697-3100, USA

We cored 13 small lakes with steep surrounding hillslopes, deep water, steep perimeter bathymetry, and inflowing streams with sandy deltas. The lakes were geographically distributed across a 20,000-km² region in Vermont and eastern New York (Supplementary Information). We used a percussion corer fitted with a piston to recover sediment cores from locations in deep water, adjacent to the stream delta foreslopes. The cores vary in length from 3.5 to 6 m, spanning 2.1 to 13.2 kyr (Fig. 1a).

We documented core stratigraphy with visual logging, magnetic susceptibility measurements, high-resolution X-radiography and loss-on-ignition². We created a composite record for each core from the results of these four primary analyses. The composite record overcomes the limitations of the individual analyses, and most accurately represents the locations and thickness of terrigenous layers. Comparison of our composite record with the results from high-resolution, whole-core grain size analysis^{11,12}, a more sensitive analysis for determining the locations of terrigenous deposits, shows good agreement, and indicates that the composite record emphasizes the most extreme sediment delivery events (Fig. 2, and Supplementary Information).

We dated the cores with 80 accelerator mass spectrometer (AMS) ¹⁴C ages of single terrestrial plant macrofossils and calibrated these dates using CALIB v4.2 (ref. 13; see also Supplementary Information). To account for the difference in sedimentation rates between the rapidly deposited terrigenous layers³ and the slow accumulation of gyttja, we excluded the terrigenous layers from each record before creating age models, and assumed constant sedimentation rates between successive radiocarbon ages². Com-

bining the age models with the composite sediment records yields a dated event record for each lake (Fig. 1a, and Supplementary Information). Using the 14 event chronologies, we created a single histogram indicating the frequency of depositional events and by inference, the most severe storms during the Holocene (Fig. 1b).

The results presented here emphasize the importance of using the sediment from multiple lakes for this type of palaeoclimate reconstruction. Storm records derived from individual lake basins² vary owing to several factors, including the geologic character of the drainage basin, local differences in antecedent soil moisture or other factors of landscape conditioning⁷, and core location. Such records may not reveal clear relationships between local events and wider climate patterns. Aggregate data derived from multiple sources are less susceptible to local environmental variability, and reveal regional landscape response to large-scale climate variability on long timescales.

The delivery of terrigenous sediment to the 13 lakes reached broad and variable maxima lasting ~1,500 yr, with the highest peaks centred at approximately 2.6, 5.8, 9.1 and 11.9 kyr before present (BP). Before European settlement of the area at ~250 yr BP, when deforestation and livestock grazing accelerated rates of hill-slope erosion (which overprints our record³), sediment delivery appears to have been increasing toward another peak. This most recent period of increased delivery began at about 600 yr BP, coincident with the beginning of the Little Ice Age (LIA)¹⁴; the earliest such period peaked during the Younger Dryas climate interval¹⁵.

Other, independent records of storminess and flooding from around the North Atlantic show maxima that correspond to those that characterize our lake records^{16–20} (Fig. 1c). For example, oscillations in Holocene glaciochemical records from central Greenland ice cores are similar to storminess records in New England lakes. In particular, the time series of sea-salt sodium (s.s. Na) in the Greenland Ice Sheet Project Two (GISP2) core, believed to be an indicator of storminess and sea spray in the atmosphere of the high-

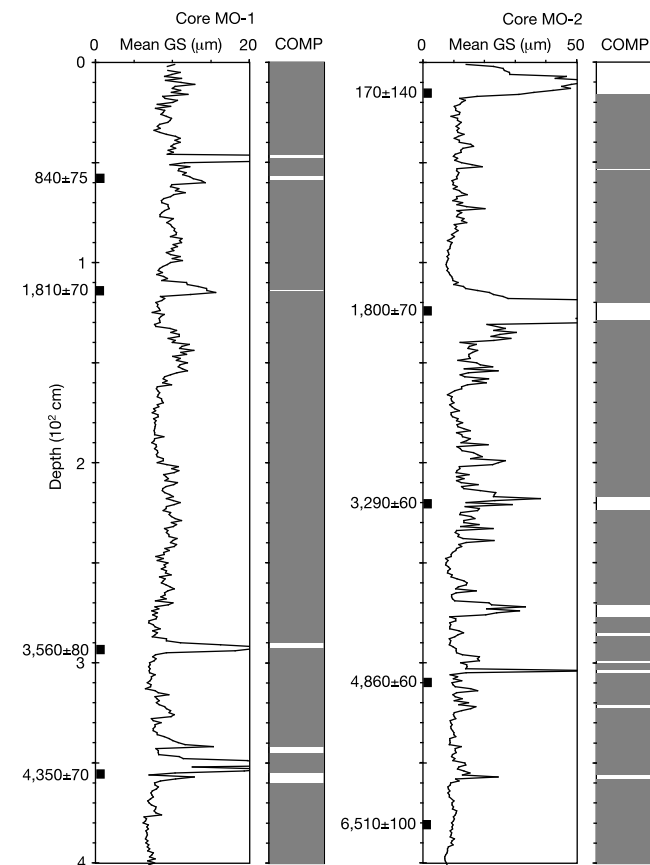


Figure 2 Results of high-resolution, whole-core grain size (GS) analysis and composite sediment record (COMP) for two Lake Morey cores. MO-1 is the distal core and MO-2 is the proximal core. Peaks in GS reflect high-energy fluvial transport events and thus the location of terrigenous layers (indicated in composite sediment record with white bands). Calibrated ¹⁴C dates (with 1σ uncertainties) shown for each core.

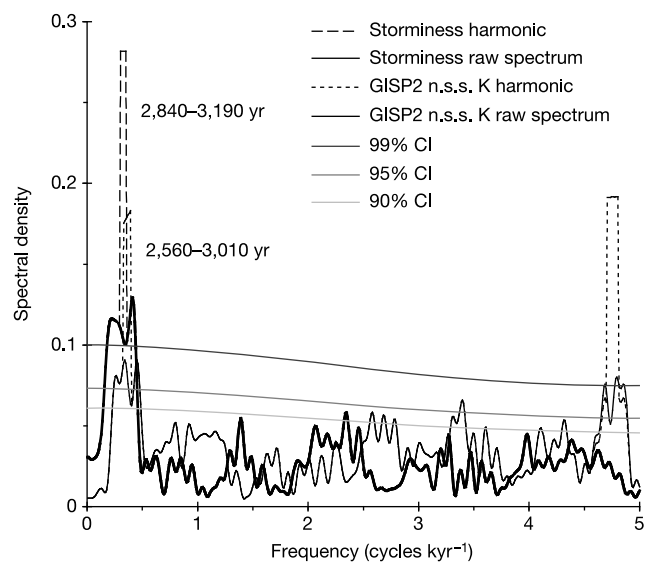


Figure 3 Multitaper spectral analysis (three tapers) of storminess and GISP2 non-sea-salt (n.s.s) K time series, interpolated with a 100-yr interval and with linear trend removed²⁹. Spectral analysis of other GISP2 aerosol deposition time series²¹ reveal power spectra similar to that of n.s.s. K. The harmonic spectrum represents the estimated significant periodic component of the raw spectrum, as measured by the Thomson variance ratio test for periodic signals (F-test). CI, confidence interval. Confidence levels are relative to red noise estimated from lag-1 autocorrelation with a median averaging filter³⁰. The n.s.s. K raw spectrum has been rescaled such that the confidence intervals apply to both power spectra.

latitude North Atlantic region^{21,22}, exhibits peaks at times similar to the maxima identified here (Fig. 1d). Other GISP2 aerosol deposition time series (notably non-sea-salt potassium, n.s.s. K) show a similar pattern (Fig. 1e). Thus, the frequency of storm-related floods in the northeastern United States appears to be related to widespread North Atlantic climate variability on millennial timescales. Lacustrine data thus support the interpretation of the GISP2 glaciochemical record in terms of large-scale atmospheric circulation changes¹⁵.

The pacing of storminess maxima derived from the various North Atlantic palaeoclimate records suggest a quasi-periodic cycle of ~3,000 yr (refs 16–21). Spectral analysis of our New England storminess time series (Fig. 3) reveals significant spectral power in a broad, double peak centred at a period of ~3,070 yr. A similar double peak exists in the power spectrum of the GISP2 time series of aerosol deposition²¹ (Figs 1d, e and 3). Monte Carlo simulations ($n = 1,000$) show that the likelihood of obtaining comparable results from a composite of arbitrary lake sediment records (produced by randomly rearranging the events identified in each individual time series) is less than 1%.

The pacing of maxima and the power spectra of the New England lake-core and Greenland ice-core time series are remarkably similar, suggesting control of both by large-scale atmospheric circulation patterns. Annually resolved Na concentrations in ice cores from North Greenland²² show a significant correlation with the characteristic patterns of sea level pressure associated with the Arctic Oscillation (AO)⁹ during the past four decades. If, as argued by O'Brien and others²¹, enhanced aerosol deposition to the Greenland ice sheet on longer timescales similarly reflects an increase in the relative strength of meridional versus zonal circulation, then increased New England storminess would coincide with a mode of atmospheric circulation most similar to the low phase of the contemporary AO, when high-latitude westerlies are weakened⁹. Conversely, times when New England storm frequency was low would coincide with the enhanced zonal flow of the high-phase AO.

Indeed, such a relationship between New England storminess and the AO has recently been confirmed for modern climate conditions by Thompson and Wallace⁹, who show that nor'easters are four times as likely to occur during low-phase than during high-phase AO. Thus, millennial-scale variability in New England storminess may reflect modulation of the preferred phase of the AO on millennial timescales. This explanation is appealing, because it makes a specific prediction that New England storminess should be at its greatest when Europe is cold (characteristic of the low-phase AO). Comparison of our results with the other climate records (Fig. 1), including European glacier fluctuations, suggests that, as predicted, intense storms in New England tend to occur more frequently during periods that are cooler than average in Europe^{15,21,23–26}.

Atmospheric modes, such as the AO and El Niño/Southern Oscillation (ENSO), account for significant fractions of modern climate variability on interannual timescales^{1,9}. In contrast to other researchers' conclusions²⁷, our results suggest that such dominant atmospheric modes may be modulated on longer timescales, and may therefore account for much of the variability in the palaeoclimate record. Changes in the AO, perhaps modulated by solar forcing²⁸, may explain a significant portion of Holocene climate variability in the North Atlantic region.

Climate models suggest that human activities, specifically the emission of atmospheric greenhouse gases, may lead to increases in the frequency of severe storms in certain regions of the Northern Hemisphere¹. However, the existence of natural variability in storminess confounds reliable detection of anthropogenic effects. During the past ~600 yr, New England storminess appears to have been increasing naturally (Fig. 1b). This rhythm in storm frequency may explain some of the recently observed increases in extreme precipitation events¹. If the pattern of millennial-scale variability

that we documented through the Holocene persists into the future, New England storminess would continue to increase for the next ~900 yr. Because climate synopses compiled from instrumental records cannot distinguish underlying natural increases in storminess from anthropogenic effects, detected increases in contemporary storminess may not be a reliable indicator of human-induced climate change. □

Methods

Layer identification

We identified significant terrigenous layers in time series by comparing peaks to the background signal. To establish appropriate background estimates, we used singular spectrum analysis (lags = 20) to identify the first principal component in each series, after setting all data points greater than 1σ from the median to equal the median. Any shifts in the original data series that were more than 1σ from the first principal component were considered to be significant. Hence, significance is conservatively measured relative to the local background, rather than the entire data series.

In the composite sediment record, most terrigenous layers were identified when two or more of the individual analytical techniques detected such layers. In some cases, layers were identified from the results of a single analysis. The composite record is similar to records produced by the individual laboratory analyses, and all resulting storminess histograms produced from these individual analyses show similar millennial-scale pacing and yield similar power spectra. We find no correlation between sediment accumulation rates and the frequency of inferred storm events ($r^2 = 0.05$).

Obtaining layer data

Our laboratory 1-cm core sampling interval produced median temporal core sampling intervals ranging from 2 yr in lakes with the highest sedimentation rates, to 27 yr in lakes with the lowest sedimentation rates, and the interval varied throughout each core owing to the effects of autocompaction and changing sedimentation rates. We used several different interpolation intervals (varying from 10 to 200 yr) to obtain evenly spaced data for spectral analysis, all of which produced similar results. The results are robust to varying resolutions and tapers, and comparable results were obtained with the multitaper, Blackman–Tukey, and maximum entropy methods, as well as periodogram analysis using the Lomb–Scargle method (which avoids the need to interpolate the data).

Testing for variability

In addition to spectral analysis, we tested our storminess time series for statistically significant variability using several other methods. First, we examined the effect of removing individual time series from our aggregate record and again performing the spectral analyses. We find that in all cases, removal of an individual record retains the spectral peak at ~3,000 yr, though in some cases with confidence levels reduced to 95% relative to red noise. Second, we created 13 new time series by rearranging the data from each according to a random uniform distribution. In this way, we created time series with the same number of events but with those events randomly distributed in time. In all cases, we find that changing one time series maintains the spectral peak at ~3,000 yr, though again in some cases with confidence levels reduced to <95% relative to red noise. If we replace two or more of the original time series with the random time series, the identified spectral peak is reduced in each case to significantly lower confidence (<90%). Third, our Monte Carlo approach tested the likelihood of obtaining a significant (at 95%) spectral peak between 2,600 and 3,400 yr using the multitaper method and 13 random time series produced as described above. Using 1,000 attempts yields a <0.005 probability of obtaining such a peak in this interval.

Received 10 June; accepted 12 September 2002; doi:10.1038/nature01132.

- Houghton, J. T., Ding, Y. & Griggs, D. J. (ed.) *Climate Change 2001: The Scientific Basis* (Cambridge Univ. Press, Cambridge, 2001).
- Brown, S. L., Bierman, P. R., Lini, A. & Southon, J. R. 10,000 yr record of extreme hydrologic events. *Geology* **28**, 335–338 (2000).
- Bierman, P. R. *et al.* Postglacial ponds and alluvial fans: Recorders of Holocene landscape history. *GSA Today* **7**, 1–8 (1997).
- Eden, D. N. & Page, M. J. Palaeoclimatic implications of a storm erosion record from late Holocene lake sediments, North Island, New Zealand. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **139**, 37–58 (1998).
- Lamoureux, S. F. Five centuries of interannual sediment yield and rainfall-induced erosion in the Canadian High Arctic recorded in lacustrine varves. *Wat. Resour. Res.* **36**, 309–318 (2000).
- Nesje, A., Dahl, S. O., Matthews, J. A. & Berrisford, M. S. A ~4500-yr record of river floods obtained from a sediment core in Lake Atnsjoen, eastern Norway. *J. Paleolimnol.* **25**, 329–342 (2001).
- Rodbell, D. T. *et al.* An ~15,000-year record of El Niño-driven alluviation in southwestern Ecuador. *Science* **283**, 516–520 (1999).
- Thorndycraft, V., Hu, Y., Oldfield, F., Crooks, P. R. J. & Appleby, P. G. Individual flood events detected in the recent sediments of the Petit Lac d'Annecy, eastern France. *Holocene* **8**, 741–746 (1998).
- Thompson, D. W. J. & Wallace, J. M. Regional climate impacts of the northern hemisphere annular mode. *Science* **293**, 85–89 (2001).
- Jennings, K. L., Bierman, P. R. & Southon, J. R. Timing and style of deposition on humid-temperate fans, Vermont, USA. *Geol. Soc. Am. Bull.* (in the press).
- Bosley, A. C., Bierman, P. R., Noren, A. J. & Galster, J. C. Identification of paleoclimatic cycles during the Holocene using grain size analysis of sediments cored from Lake Morey in Fairlee, VT. *Geol. Soc. Am. Abstr. Prog.* **33**, 15 (2001).
- Conlan, A. *Vermont Geological Society Spring Meeting* (Norwich Univ., Norwich, VT, 2001).

13. Stuiver, M. & Reimer, P. J. Extended ^{14}C data base and revised CALIB 3.0 ^{14}C age calibration program. *Radiocarbon* **35**, 215–230 (1993).
14. Grove, J. M. *The Little Ice Age* (Methuen, London, 1988).
15. Mayewski, P. A. *et al.* Changes in atmospheric circulation and ocean ice cover over the North Atlantic during the last 41,000 years. *Science* **263**, 1747–1751 (1994).
16. Brown, P., Kennett, J. P. & Ingram, B. L. Marine evidence for episodic Holocene megafloods in North America and the northern Gulf of Mexico. *Paleoceanography* **14**, 498–510 (1999).
17. Knox, J. C. Sensitivity of modern and Holocene floods to climate change. *Quat. Sci. Rev.* **19**, 439–457 (1999).
18. Lamb, H. H. Variation and changes in the wind and ocean circulation: the Little Ice Age in the northeast Atlantic. *Quat. Res.* **11**, 1–20 (1979).
19. Liu, K. b. & Fearn, M. L. Reconstruction of prehistoric landfall frequencies of catastrophic hurricanes in northwestern Florida from lake sediment records. *Quat. Res.* **54**, 238–245 (2000).
20. Zong, Y. & Tooley, M. J. Evidence of mid-Holocene storm-surge deposits from Morecambe Bay, northwest England: A biostratigraphical approach. *Quat. Int.* **55**, 43–50 (1999).
21. O'Brien, S. R. *et al.* Complexity of Holocene climate as reconstructed from a Greenland ice core. *Science* **270**, 1962–1964 (1995).
22. Fischer, H. Imprint of large-scale atmospheric transport patterns on sea-level records in northern Greenland ice cores. *J. Geophys. Res.* **106**, 23977–23984 (2001).
23. Hormes, A., Müller, B. U. & Schlüchter, C. The Alps with little ice: evidence for eight Holocene phases of reduced glacier extent in the Central Swiss Alps. *Holocene* **11**, 255–265 (2001).
24. Karlén, W. & Kylenstierna, J. On solar forcing of Holocene climate: evidence from Scandinavia. *Holocene* **6**, 359–365 (1996).
25. Boulton, G. S., *et al.* *Glacier Fluctuations during the Holocene* (eds Frenzel, B., Boulton, G. S., Glaser, B. & Hückriede, U.) 5–23 (Fischer, Stuttgart, 1997).
26. Matthews, J. A., Dahl, S. O., Nesje, A., Berrisford, M. & Andersson, C. Holocene glacier variations in central Jotunheimen, southern Norway based on distal glaciolacustrine sediment cores. *Quat. Sci. Rev.* **19**, 1625–1647 (2000).
27. Bond, G. *et al.* Persistent solar influence on North Atlantic climate during the Holocene. *Science* **294**, 2130–2136 (2001).
28. Shindell, D. T., Schmidt, G. A., Mann, M. E., Rind, D. & Waple, A. Solar forcing of regional climate change during the Maunder minimum. *Science* **294**, 2149–2152 (2001).
29. Thomson, D. J. Spectrum estimation and harmonic analysis. *Proc. IEEE* **70**, 1055–1096 (1982).
30. Mann, M. E. & Lees, J. M. Robust estimation of background noise and signal detection in climatic time series. *Clim. Change* **33**, 409–445 (1996).

Supplementary Information accompanies the paper on Nature's website
(<http://www.nature.com/nature>).

Acknowledgements We thank J. Galster for field and laboratory assistance, K. Jennings for field assistance, A. Bosley and A. Conlan for grain size analysis, S. Brown for laboratory training, P.T. Davis for coring device design and training, and J. M. Wallace and H. Fischer for comments and discussion. This work was supported by the US National Science Foundation (P.R.B.).

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to A.J.N.
(e-mail: anders.noren@alumni.carleton.edu).

The strength of $\text{Mg}_{0.9}\text{Fe}_{0.1}\text{SiO}_3$ perovskite at high pressure and temperature

Jiuhua Chen, Donald J. Weidner & Michael T. Vaughan

Mineral Physics Institute and Department of Geosciences, State University of New York at Stony Brook, Stony Brook, New York 11794-2100, USA

The Earth's lower mantle consists mainly of $(\text{Mg,Fe})\text{SiO}_3$ perovskite and $(\text{Mg,Fe})\text{O}$ magnesio-wüstite, with the perovskite taking up at least 70 per cent of the total volume¹. Although the rheology of olivine, the dominant upper-mantle mineral, has been extensively studied, knowledge about the rheological behaviour of perovskite is limited. Seismological studies indicate that slabs of subducting oceanic lithosphere are often deflected horizontally at the perovskite-forming depth, and changes in the Earth's shape and gravity field during glacial rebound indicate that viscosity increases in the lower part of the mantle. The rheological properties of the perovskite may be important in governing these phenomena. But $(\text{Mg,Fe})\text{SiO}_3$ perovskite is not

stable at high temperatures under ambient pressure, and therefore mechanical tests on $(\text{Mg,Fe})\text{SiO}_3$ perovskite are difficult. Most rheological studies of perovskite have been performed on analogous materials^{2–7}, and the experimental data on $(\text{Mg,Fe})\text{SiO}_3$ perovskite are limited to strength measurements at room temperature in a diamond-anvil cell⁸ and microhardness tests at ambient conditions⁹. Here we report results of strength and stress relaxation measurements of $(\text{Mg}_{0.9}\text{Fe}_{0.1})\text{SiO}_3$ perovskite at high pressure and temperature. Compared with the transition-zone mineral ringwoodite¹⁰ at the same pressure and temperature, we found that perovskite is weaker at room temperature, which is consistent with a previous diamond-anvil-cell experiment⁸, but that perovskite is stronger than ringwoodite at high temperature.

The perovskite sample was synthesized at 26 GPa and 2,100 K from a glass of composition $(\text{Mg}_{0.9}\text{Fe}_{0.1})\text{SiO}_3$. X-ray diffraction confirmed the pure perovskite phase. *In situ* stress measurements were performed using a large-volume press (SAM85) at the X17B beamline of the National Synchrotron Light Source¹¹. Stress in the sample was measured by monitoring the broadening of diffraction peaks from the sample at high pressure and temperature. The experiment is described briefly below; details of the methodology, along with previous studies on mantle-related materials, can be found elsewhere^{10,12,13}.

The synthesized sample was powdered at liquid-nitrogen temperature to prevent amorphization, and loaded into a high-pressure cell together with a layer of NaCl as an internal pressure calibrant. The experiment was conducted by first increasing the pressure to 20 GPa, and then increasing the temperature. At each temperature, energy-dispersive X-ray diffraction patterns were collected as a function of time to characterize the stress relaxation. The sample pressure was monitored by comparing the compression of NaCl to the Decker scale¹⁴, and the sample temperature was measured by a W_{97}Re_3 – $\text{W}_{75}\text{Re}_{25}$ thermocouple. In all the X-ray diffraction patterns from the sample at the pressures and temperatures reported here, no back transformation was observed. When temperature was later increased above 1,273 K, the diffraction pattern showed a back transformation in the sample because the experiment was carried out at a pressure lower than the perovskite stability field. Our stress measurements are therefore limited to the data collected up to 1,073 K.

Heterogeneous deviatoric stress at each grain of the powdered sample, together with very small grain sizes, gives rise to diffraction

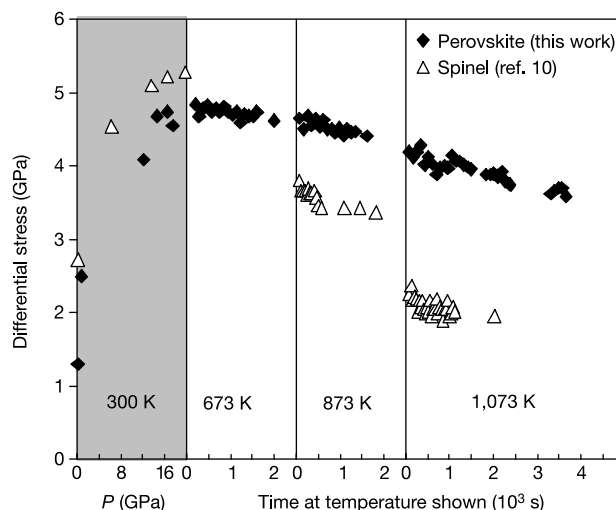


Figure 1 Differential stress supported by $\text{Mg}_{0.9}\text{Fe}_{0.1}\text{SiO}_3$ perovskite and Mg_2SiO_4 spinel as a function of time at several constant temperatures under ~20-GPa confining pressure. Filled diamonds, perovskite (this work); open triangles, spinel (from previous study¹⁰). Data in the shaded area correspond to stresses of the sample during loading.

peak broadening. In the case of energy dispersive X-ray diffraction, stress-induced peak broadening is proportional to the photon energy of the X-ray, whereas the broadening caused by small grain size is energy independent¹⁵, that is:

$$\text{FWHM}^2 = (2\varepsilon E)^2 + (Kh c/2P \sin \Theta \text{kern} + 1_0)^2$$

where FWHM represents the sample contribution to the peak broadening (full-width at half-maximum of a gaussian profile), 2ε is the differential strain (which is equal to the differential stress multiplied by the aggregate Young's modulus (442.4 GPa is used for the perovskite)), E is the X-ray photon energy, K is the Scherrer constant, h is Planck's constant, c is the velocity of light, P is the average grain size, and $2\theta_0$ is the fixed scattering angle. The results reported here are derived by deconvoluting the peak broadening of the sample diffraction patterns from observed signal^{10,13}.

Figure 1 shows the measured differential stress supported by the perovskite sample as a function of time. The grey panel shows data obtained as pressure was first increased at 300 K; the other panels show data obtained at several other constant temperatures at ~20-GPa confining pressure. For comparison, the data for γ -Mg₂SiO₄ spinel (ringwoodite)¹⁰ are superposed on the plot. The stresses in both the perovskite and the ringwoodite samples increase with load pressure, and a downward curvature of the stress versus load pressure indicates plastic yielding of the sample. At room temperature, the strength of the (Mg_{0.9}Fe_{0.1})SiO₃ perovskite is about 0.5 GPa lower than that of the ringwoodite. However, the perovskite becomes much stronger relative to the ringwoodite at higher temperatures. At 873 K, the strength of the (Mg_{0.9}Fe_{0.1})SiO₃ perovskite is about 1 GPa higher than that of the ringwoodite, and the difference increases to 2 GPa at 1,073 K. We find that the strength of perovskite at room temperature is slightly less than that of ringwoodite; this result is qualitatively consistent with that of ref. 8. However, at high temperature, the perovskite is definitely stronger than the ringwoodite.

All of these measurements are in a regime of high stress and relatively low temperature. At these conditions, there are a few possible deformation mechanisms that may be operative. Assuming a constant total strain (elastic + plastic) during the relaxation process¹³, we derived the plastic strain rate from the time-resolved data at temperatures of 673 K, 873 K and 1,073 K in order to explore the possible operative mechanism. Dislocation glide is often the dominant deformation mechanism under high stress and low

temperature¹⁶. The strain rate is generally described by the Arrhenius equation:

$$\dot{\varepsilon} = A_1 \left(\frac{\sigma}{\mu} \right)^2 \exp(-Q(\sigma)/RT) \quad (1)$$

where $\dot{\varepsilon}$ is the strain rate, A_1 a constant, μ the shear modulus, R the gas constant, T the temperature, and Q the activation energy (which is a function of the stress σ). We could not find an adequate set of parameters to fit the experimental data at all three temperatures. When we divided the data into two sets—(1) 673 K and 873 K, and (2) 873 K and 1,073 K—and expressed the stress effect on Q as $Q(1 - \sigma/\tau)$, where τ represents the Peierls stress (that is, lattice friction stress at temperature of 0 K), we found that both τ and Q change as a function of temperature (5.8 GPa and 183 kJ mol⁻¹ for case (1); 7.6 GPa and 236 kJ mol⁻¹ for case (2)).

Another possible deformation mechanism is power-law creep by climb-plus-glide, which is characterized by:

$$\dot{\varepsilon} = A_2 \frac{D_v}{RT} \left(\frac{\sigma}{\mu} \right)^n \quad (2)$$

where A_2 is a constant, and D_v the lattice diffusion coefficient. The exponent n normally varies from 3 to 10. However, the stress relaxation in the perovskite is so slow that the derived exponent n is about one order of magnitude higher than the typical values. Therefore this power-law creep is also unlikely to be the operative mechanism.

Thus it is likely that among the competing deformation mechanisms in the perovskite system, some other relaxation process is weaker than dislocation glide and power-law creep, and hence controls the strength in this temperature region. One possibility is twinning. Twinning domains are easily generated in (Mg,Fe)SiO₃ perovskite on (110) and (112) planes because of its ferroelastic character¹⁷. The (110) twinning results in an interchange of the a and b axes between adjacent twin domains^{18,19}, and can be induced by differential stress. Under a cylindrical stress field, as in our experiment¹¹, the twinning on the (110) plane during compression gives rise to an increase of relative diffraction intensity ($I_{(200)}/I_{(020)}$) in the perovskite structure²⁰. Because of the clear observation of such features in our experiment (Fig. 2) and the support from the previous related studies²⁰, we suggest that mechanical twinning may be a favourable candidate for an operative deformation mechanism in the (Mg,Fe)SiO₃ perovskite under the high-stress and low-temperature regime. However, this suggestion still needs to be confirmed by some more-direct evidence, although the preferred mechanical twinning mechanism with respect to dislocation glide at low temperature is also observed in other minerals²¹. Although the strain rate for the twinning mechanism can be estimated by an equation similar to equation (1), it has been demonstrated that both effective activation energy (Q) and twinning nucleation stress (τ) are temperature dependent¹⁶. This is consistent with the characteristics of our data.

On the other hand, another mechanism—dynamic recrystallization—is known to produce strength that may have a weak temperature dependence in some cases^{22,23}. This process might also seem to be a potential candidate for the operative deformation mechanism because of the observed weak temperature dependence of the perovskite strength (Fig. 1). But dynamic recrystallization has been documented to take place mostly at high temperature ($T > 0.6T_m$)¹⁶, T_m being the melting temperature and our data are limited to the temperatures $< 0.5T_m$.

Although the operative deformation mechanism is not known for sure, our experiments show that perovskite is very strong (stronger than ringwoodite) at these conditions. If twinning is the mechanism, it could only operate for small strains and would not be viable once the entire sample had obtained the minimum-energy twin state: a stronger deformation mechanism might then become

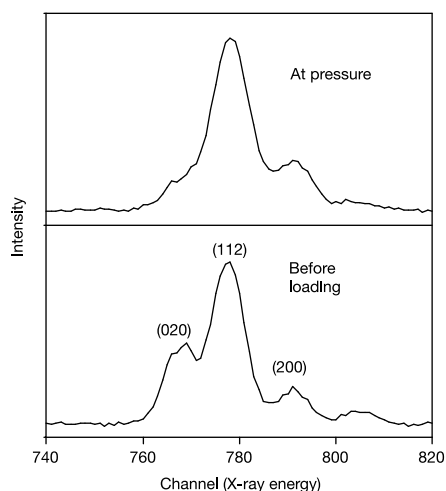


Figure 2 Change in the relative intensities of the (020) and (200) diffraction peaks of Mg_{0.9}Fe_{0.1}SiO₃ perovskite owing to compression. The data from the compressed sample indicate an increase in the population of (200)-oriented grains (parallel to the loading axis) relative to (020)-oriented grains. This is evidence for a twinning process that converts the b axis to the shorter a axis under uniaxial loading (see text).

operative. The current result represents the lower limits of strength of the perovskite at these conditions.

We note that the pressure, temperature and stress levels in our experiment are not the same as the conditions in the Earth's lower mantle, where the temperature is probably 1,000 K higher and the stress level is about two orders of magnitude lower. Therefore the deformation mechanism probably falls into a different regime, and the current result can not be directly applied to the lower mantle. But this study reveals the rheological behaviour of (Mg,Fe)SiO₃ perovskite at high pressure and temperature, and suggests a revision of expectations that are based on room-temperature results⁸. It is probably more reasonable to predict the rheological behaviour of perovskite in the mantle from the present high-pressure and high-temperature experimental result, than from high-pressure but room-temperature data. Under the experimental conditions, using the data for both perovskite and ringwoodite (Fig. 1), we estimated the viscosity ($\eta = \sigma/\dot{\epsilon}$) of perovskite to be one order of magnitude higher than that of ringwoodite at 1,073 K. Once again, as the deformation in the Earth's lower mantle is likely to be dominated by diffusion creep, a more accurate comparison should be made on the basis of the diffusion creep data when they become available.

On the other hand, 1,073 K is a very reasonable temperature to be expected in subducted slabs at the bottom of the transition zone. Our experiment clearly shows that at this temperature perovskite is definitely much stronger than ringwoodite. This fact helps to explain why some subducted slabs deflect at top of the lower mantle, as the perovskite may create a mechanical barrier to subduction of the weaker ringwoodite²⁴. Ultimately, the determination of the relative strengths of the different media must also consider other characteristics, such as grain-size effects as discussed by Karato *et al.*²⁵. These authors showed that coarsening of the grain size increases the slab strength, and therefore a coarse-grained slab can penetrate into the lower mantle.

The mechanism of deep-focus earthquakes along subducted slabs has been a long-standing question, because the traditional brittle failure model is not likely to be operative in the deep mantle. With the progress in understanding the rheological properties of minerals, the early model of a thermal runaway instability^{26,27}—when deformation of a material occurs so rapidly that heat is accumulated in the region of high strain—has regained special attention^{25,28}. Using existing rheology data, Karato *et al.*²⁵ have modelled the rheological structure of subducted slabs in a fashion that satisfies the depth variation of seismicity down to the bottom of the transition zone. Lacking the rheology data for perovskite, these authors inferred that the clear shut-off of seismicity at the top of the lower mantle is probably attributable to loss of strength of the slabs in this depth region, owing to superplasticity and/or the absence of strong forces. In fact, following the thermal runaway instability model, thermal weakening (for example, the temperature-sensitive strength of olivine and ringwoodite¹⁰) enhances the positive feedback between deformation-induced heating and deformation. If the subducted slabs in the lower mantle consist of cold perovskite which does not weaken significantly with temperature as we observed (Fig. 1), the thermal runaway instability may not be promoted, with the result that earthquakes would not occur in the perovskite portion of the slab. □

Received 27 March; accepted 12 September 2002; doi:10.1038/nature01130.

1. Anderson, D. L. & Bass, J. D. Transition region of the Earth's upper mantle. *Nature* **320**, 321–328 (1986).
2. Poirier, J. P., Peyronneau, J., Gesland, J. Y. & Brebec, G. Viscosity and conductivity of the lower mantle: an experimental study on a MgSiO₃ perovskite analogue, KZnF₃. *Phys. Earth Planet. Inter.* **32**, 273–287 (1983).
3. Doukhan, N. & Doukhan, J. C. Dislocation in perovskite BaTiO₃ and CaTiO₃. *Phys. Chem. Miner.* **13**, 403–410 (1986).
4. Karato, S.-i. Plasticity-crystal structure systematics in dense oxides and its implications for the creep strength of the Earth's deep interior: a preliminary result. *Phys. Earth Planet. Inter.* **55**, 234–240 (1989).
5. Karato, S.-i. & Li, P. Diffusion creep in perovskite: implications for the rheology of the lower mantle. *Science* **255**, 1238–1240 (1992).

6. Li, P., Karato, S.-i. & Wang, Z. High-temperature creep in fine-grained polycrystalline CaTiO₃, an analogue material of (Mg,Fe)SiO₃ perovskite. *Phys. Earth Planet. Inter.* **95**, 19–36 (1996).
7. Wright, K., Price, G. D. & Poirier, J. P. High-temperature creep of the perovskite CaTiO₃ and NaNbO₃. *Phys. Earth Planet. Inter.* **74**, 9–22 (1992).
8. Meade, C. & Jeanloz, R. The strength of mantle silicates at high pressures and room temperature: implications for the viscosity of the mantle. *Nature* **348**, 533–535 (1990).
9. Karato, S.-i., Fujino, K. & Ito, E. Plasticity of MgSiO₃ perovskite: the results of microhardness tests on single crystals. *Geophys. Res. Lett.* **17**, 13–16 (1990).
10. Chen, J., Inoue, T., Weidner, D., Wu, Y. & Vaughan, M. Strength and water weakening of mantle minerals, olivine, wadsleyite and ringwoodite. *Geophys. Res. Lett.* **25**, 575–578 1103–1104 (1998).
11. Weidner, D. J., *et al.* *High-pressure Research: Application to Earth and Planetary Sciences* (eds Syono, Y. & Manghnani, M. H.) 13–17 (Terra Scientific/AGU, Tokyo/Washington DC, 1992).
12. Weidner, D. J., Wang, Y. & Vaughan, M. T. Yield strength at high pressure and temperature. *Geophys. Res. Lett.* **21**, 753–756 (1994).
13. Weidner, D. J., Wang, Y., Chen, G., Ando, J. & Vaughan, M. T. *Properties of Earth and Planetary Materials at High Pressure and Temperature* (eds Manghnani, M. H. & Yagi, T.) 473–482 (Geophysical Monograph, AGU, Washington DC, 1998).
14. Decker, D. L. High-pressure equation of state for NaCl, KCl and CsCl. *J. Appl. Phys.* **42**, 3239–3244 (1971).
15. Gerward, L., Morup, S. & Topsoe, H. Particle size and strain broadening in energy-dispersive x-ray powder patterns. *J. Appl. Phys.* **47**, 822–825 (1976).
16. Frost, H. J. & Ashby, M. F. *Deformation-Mechanism Maps* (Pergamon, Oxford, 1982).
17. Wang, Y., Guyot, F., Yeganeh-Haeri, A. & Lieberman, R. C. Twinning in MgSiO₃ perovskite. *Science* **248**, 468–471 (1990).
18. Sapriel, J. Domain-wall orientations in ferroelastics. *Phys. Rev. B* **12**, 5128–5140 (1975).
19. Toledano, J. C. & Toledano, P. Order parameter symmetries and free-energy expansions for purely ferroelastic transitions. *Phys. Rev. B* **21**, 1139–1172 (1980).
20. Zhao, Y. *et al.* High-pressure crystal chemistry of neighborite, NaMgF₃: An angle-dispersive diffraction study using monochromatic synchrotron x-radiation. *Am. Mineral.* **79**, 615–621 (1994).
21. Avé Lallemant, H. G. Experimental deformation of diopside and websterite. *Tectonophysics* **48**, 1–27 (1978).
22. Derby, B. *Deformation Processes in Minerals, Ceramics and Rocks* (eds Barber, D. J. & Meredith, P. G.) 354–364 (Unwin Hyman, London, 1990).
23. Shimizu, I. Stress and temperature dependence of recrystallized grain size: A subgrain misorientation model. *Geophys. Res. Lett.* **25**, 4237–4240 (1998).
24. Gurnis, M. & Hager, B. H. Controls of the structure of subducted slabs. *Nature* **335**, 317–321 (1988).
25. Karato, S.-i., Riedel, M. R. & Yuen, D. A. Rheological structure and deformation of subducted slabs in the mantle transition zone; implications for mantle circulation and deep earthquakes. *Physics Earth Planet. Inter.* **127**, 83–108 (2001).
26. Griggs, D. T. & Baker, D. W. *Properties of Matter Under Unusual Conditions* (eds Mark, H. & Fernbach, S.) 23–42 (Wiley/Interscience, New York, 1969).
27. Ogawa, M. Shear instability in a viscoelastic material as the cause of deep focus earthquakes. *J. Geophys. Res.* **92**, 13801–13810 (1987).
28. Weidner, D. J. *et al.* Subduction zone rheology. *Phys. Earth Planet. Inter.* **127**, 67–81 (2001).

Acknowledgements This work was carried out at the X17B beamline of the National Synchrotron Light Source. We thank Z. Zhang, J. B. Hastings and D. P. Siddons for technical support at the beamline, and J. Zhang for help in sample preparation. This work was supported by the US National Science Foundation.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to J.C. (e-mail: jiuhsia.chen@sunysb.edu).

Contemporary fisherian life-history evolution in small salmonid populations

Mikko T. Koskinen*, Thron O. Haugen† & Craig R. Primmer*

* Department of Ecology and Systematics, Division of Population Biology, PO Box 65, Viikinkaari 1, FIN-00014, University of Helsinki, Finland

† Department of Biology, Division of Zoology, PO Box 1050, Blindern, N-0316, Oslo, Norway

The relative importance of natural selection¹ and random drift² in phenotypic evolution has been discussed since the introduction of the first population genetic models^{3–5}. The empirical evidence used to evaluate the evolutionary theories of Fisher¹ and Wright² remains obscure because formal tests for neutral divergence^{6–8} or sensitive attempts to separate the effects of

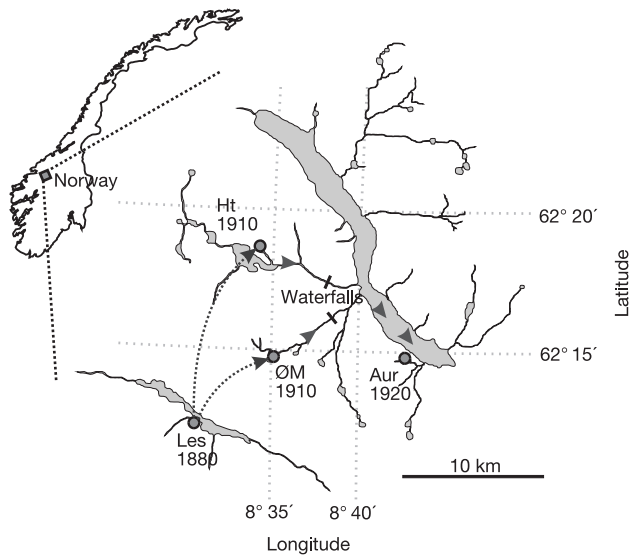


Figure 1 Map presenting the study region and location of the *T. thymallus* populations. The population in Lake Lesjaskogsvatn (Les) was founded in 1880 and used for introducing a small number of grayling into lakes Hårtjønn (Ht) and Øvre Mærrabottvatn (ØM) in 1910. These populations were the sources for colonization of Aursjøen (Aur), where grayling were first observed in 1920. All grayling in the system originate from these introductions. Land or impassable waterfalls prevent all migration between Les–Ht, Les–ØM, Les–Aur and Ht–ØM.

selection and drift are scarce, subject to error, and have not been interpreted in the light of well-known population demography. We combined quantitative genetic and microsatellite DNA analyses to investigate the determinants of contemporary life-history evolution in isolated populations of grayling (*Thymallus thymallus*, Salmonidae) that originated from a common source 80–120 years ago. Here we show that natural selection was the dominant diversifying agent in the evolution of the quantitative traits. However, the populations were founded by a small number of individuals, exhibit very low microsatellite-based effective sizes and show genetic imprints of severe ‘bottlenecks’; which

are conditions often suggested to constrain selection and favour drift^{6,8,9}. This study demonstrates a very clear case of fisherian evolution in small natural populations across a contemporary timescale.

The fisherian view of evolution holds that phenotypic differentiation results primarily from positive natural selection¹. In contrast, the influential theories developed by Wright² and evolutionists working after Wright^{6,8,9} emphasize the potential importance of random drift in effectively small populations. Evidence from species experiencing changing environmental conditions suggest that directional selection can be powerful in nature^{10–14}. However, the evolutionary theories warrant further study for several reasons.

First, formal tests for the null-hypothesis of neutral evolution^{6–8} have rarely been applied to wild organisms^{8,15}. This is because the tests require knowledge of mutational input, divergence time (*t*) and effective population size (*N_e*), which are normally unknown or subject to considerable error⁵. Second, despite the published examples of natural selection^{10–14}, sensitive attempts to evaluate the potential effect of drift in phenotypic evolution remain scarce. A promising indirect method is to compare genetic differentiation based on quantitative traits (*Q_{ST}*)¹⁶ and DNA markers affected by drift (*F_{ST}*)¹⁷. Third, studies applying the neutrality tests^{8,15} or attempting to separate the causes of evolution¹⁶ were performed for populations that diverged thousands of generations ago, leaving much room for error owing to migration from unknown sources, differing mutational contribution on coding genes and neutral markers, or convergence of *F_{ST}* on its maximal value⁵. Thus, although these two important methods of evolutionary inference are most useful across contemporary timescales, such applications have not been reported. Finally, studies linking the relative importance of selection and drift with well-known population demography are nonexistent. This is a major caveat for evaluation of the evolutionary theories because many models consider *N_e* as the key parameter expected to influence the power of the alternative diversifying agents^{6,8,9}.

Here we present an investigation in an exceptional natural model system that provides a sensitive framework for assessing the relative importance of selection and drift on quantitative trait divergence. Importantly, this study effectively avoids the above-mentioned sources of bias and involves populations with well-known demographic histories, providing an unequivocal evaluation of the

Table 1 Evidence for natural selection

Trait	Pairwise comparison	<i>h</i> ²	<i>s</i> ¹	σ^2_{GW}	σ^2_{GB}	<i>F</i> _{1,∞}	<i>P</i>	<i>N_e</i> (sign)
Length at termination (mm)	Les–Ht	0.10	0.95	0.08	1.47	384	***	0.25
	Les–Aur	0.06	0.58	0.18	0.03	10.7	*	22.6
	Ht–Aur	0.06	1.07	0.27	0.72	134	***	0.93
	Les–Ht	0.05	1.73	0.17	2.40	647	***	0.14
Yolk-sac volume (mm ³)	Les–Aur	0.03	3.24	0.44	5.55	1.78 × 10 ³	***	0.12
	Ht–Aur	0.03	2.07	0.46	0.45	99.3	***	1.54
	Les–Ht	0.18	0.41	6.57 × 10 ^{−5}	6.55 × 10 ^{−4}	111	***	0.84
	Les–Aur	0.22	0.21	7.23 × 10 ^{−5}	9.72 × 10 ^{−5}	28.2	***	7.54
Growth rate (mm/Δ <i>D</i>)	Ht–Aur	0.22	0.42	1.73 × 10 ^{−5}	2.22 × 10 ^{−4}	194	***	0.79
	Les–Ht	0.09	0.66	0.08	0.23	65.7	***	1.42
	Les–Aur	0.13	0.32	0.12	0.08	22.0	***	8.84
	Ht–Aur	0.13	0.38	0.27	0.10	9.43	*	16.6
Incubation time (days)	Les–Ht	0.22	0.26	0.26	0.50	17.4	***	5.33
	Les–Aur	0.44	0.04	0.30	1.00 × 10 ⁸	3.45 × 10 ^{−7}	>0.99	6.16 × 10 ⁸
	Ht–Aur	0.44	0.13	0.24	0.58	18.0	**	8.49
	Les–Ht	0.11	0.72	0.02	2.50 × 10 ^{−3}	2.50	0.12	37.1
Swim-up length (mm)	Les–Aur	0.32	0.03	0.04	3.13 × 10 ^{−4}	0.12	0.73	1.73 × 10 ³
	Ht–Aur	0.32	0.22	0.11	9.28 × 10 ^{−10}	8.32 × 10 ^{−8}	>0.99	1.83 × 10 ⁹
	Les–Ht	0.15	0.38	0.51	3.62 × 10 ^{−9}	9.53 × 10 ^{−8}	>0.99	9.76 × 10 ⁸
	Les–Aur	0.08	0.53	0.27	2.37 × 10 ^{−9}	4.84 × 10 ^{−7}	>0.99	4.40 × 10 ⁸
Hatching length (mm)	Ht–Aur	0.08	1.11	0.03	6.09 × 10 ^{−4}	0.70	0.40	218

*h*² is the narrow-sense heritability for the youngest population in each comparison. The standardized selection differentials (*s*¹) indicating the strength of directional selection (in s.d. units per generation), were obtained from the ratio of haldane estimates¹⁸ and the corresponding *h*² estimates²⁴. σ^2_{GW} , Additive genetic variance within populations; σ^2_{GB} , Additive genetic variance between populations. *F* = (*N_e* σ^2_{GB})/(*h*² σ^2_{GW} *t*) (refs 6, 7), where *N_e* is the maximum likelihood estimate of effective population size (Table 2; Supplementary Information), and *t* equals 11.8, 11.8 and 12.3 generations between Les–Ht, Les–Aur and Ht–Aur, respectively. The *P*-values have been adjusted for multiple significance tests using a Bonferroni correction. ***, *P* < 0.0001; **, 0.0001 ≤ *P* < 0.001; *, 0.001 ≤ *P* < 0.01. *N_e* (sign) is the level of effective population size required to produce an *F*-value leading to rejection of the null-hypothesis at the *P* ≤ 0.05 significance level (*F*_{1,∞} = 3.84).

determinants of phenotypic evolution.

Grayling inhabiting Lake Lesjaskogsvatn (Les), Norway, originate from a single human-mediated introduction in 1880 (Fig. 1). Thirty years later, a local fisherman carried a small number of individuals from Lesjaskogsvatn into two nearby mountain lakes, Hårrtjønn (Ht) and Øvre Mærrabottvatn (ØM). The exact number of individuals in the fisherman's bucket remains unknown, but was very small because Ht and ØM are separated from Les by approximately five hours of demanding uphill hiking, making it impossible for the fisherman to transfer large volumes of water (J. Nordsletten, personal communication). The founders of Ht and ØM (or their offspring) subsequently dispersed downstream into Aursjøen (Aur), where grayling were first observed in 1920 (Fig. 1). All *T. thymallus* in the system (Fig. 1) are well known to originate from these stocking and dispersal events (J. Nordsletten, personal communication). When the divergence times of the populations are combined with their generation intervals in Les, Ht, ØM and Aur (4.0–7.5 yr, as estimated using the 'life table method' that accounts for overlapping generations¹⁸), it follows that the populations coalesce to a common ancestor 11.8–22.0 generations ago. With so few generations, diversifying mutations are negligible causes of phenotypic differentiation (see discussion in ref. 19). Further, migration to three of the populations has been impossible since their founding, either due to barriers formed by land (Les; Fig. 1) or impassable waterfalls (Ht and ØM; Fig. 1).

The null-hypothesis of neutral evolution was tested^{6,7} for divergence of seven life-history traits obtained from a 'common-garden' experiment. The null-hypothesis was rejected for the majority of the quantitative traits with a very high level of confidence ($F_{1,\infty}$; $P < 0.0001$; Table 1). Evidence against divergence at a neutral rate was compelling in all of the pairwise population comparisons for length at termination, yolk-sac volume, growth rate and survival (Table 1). For these traits, the additive genetic interpopulation differences were so large that unrealistically small N_e estimates would have been necessary for evolution to proceed due to drift

alone (Table 1). For instance, neutral divergence of yolk-sac volume would have required an average N_e of only 0.12–1.54 individuals (mean across populations, 0.60) throughout the evolutionary time periods of the grayling populations (Table 1). Such small theoretical estimates exclude drift as being the dominant evolutionary process. Strength of selection was estimated also using standardized selection differentials (s'). These estimates revealed that most traits had been under very strong directional selection pressure: the mean selection differentials across traits (among populations) were 0.71–0.77 (Table 1), which are above the 96th percentile of the estimates normally observed in natural populations²⁰.

We compared among-population differences based on additive genetic variances from the common-garden experiment (Q_{ST} ¹⁶) and analogous measures based on microsatellite DNA from analysis of 17 loci (F_{ST} ¹⁷; microsatellite allele frequencies are available in the Supplementary Information). This is a well-established indirect method for separating the causes of adaptive divergence¹⁶. Well in line with the neutrality test results, the quantitative trait divergence often greatly exceeded the microsatellite divergence (Fig. 2). Hence, evolution of many of the traits was driven by divergent natural selection instead of random drift. For instance, the pairwise Q_{ST} estimates for growth rate (0.36–0.88) were 4–10 times higher than the corresponding F_{ST} estimates (Fig. 2). Importantly, our Q_{ST}/F_{ST} comparisons are not biased by phenotypic plasticity (genetic variances were obtained from a common-garden experiment), maternal effects (sire components of additive genetic variances were used for the Q_{ST} estimation¹⁹), or mutations (owing to the short coalescence times), factors which have commonly been a potential source for error⁵. We have assumed that physical linkages between the microsatellites and genes determining the life-history evolution are not biasing the Q_{ST}/F_{ST} comparisons. Such linkages, however, would only make our conclusions regarding the dominating effect of selection conservative: divergent selective sweeps to microsatellite allele frequencies would increase the F_{ST} estimates.

We note that all of the life-history traits included in this study are very important for fitness of *T. thymallus*²¹ and that many of the traits evolved primarily owing to directional selection (Table 1; Fig. 2). When offspring of the populations were reared in the three distinct July–August water temperatures naturally occurring within Lesjaskogsvatn, Hårrtjønn and Aursjøen, they performed best in the experimental temperatures that corresponded precisely with the natural conditions²¹. For instance, grayling originating from the naturally warm Lake Hårrtjønn exhibited the highest survival rate in warm experimental conditions (and lowest survival in cold conditions), whereas individuals from the naturally cold Lake Aursjøen exhibited highest survival in a cold environment (and lowest survival in warm conditions). Accordingly, grayling originating from Lesjaskogsvatn, a lake with medium natural temperatures, exhibited highest survival rates in medium temperature conditions²¹. This population $\times$ temperature interaction effect was highly significant ($F_{4,148} = 6.53$, $P < 0.0001$)²¹. It therefore seems very likely that selection acted in favour of local adaptation to the specific temperatures that the populations experience in nature.

The fisherian view of evolution¹ has often been challenged by

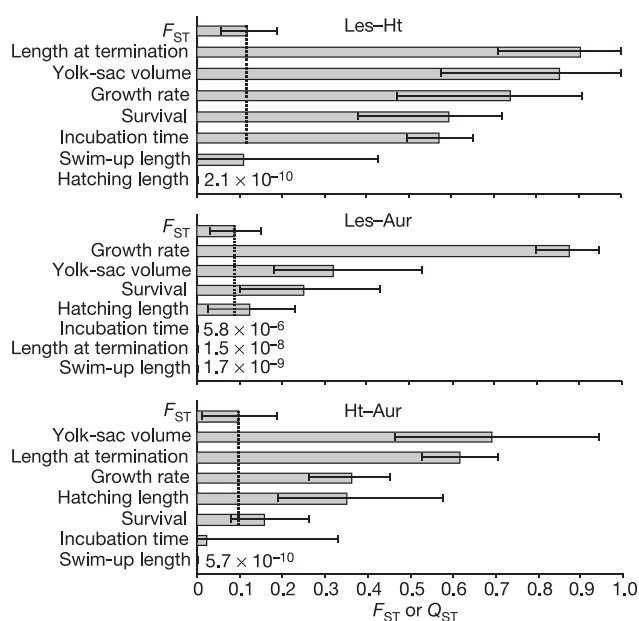


Figure 2 Among-population genetic differences based on additive genetic variances of seven life-history traits (Q_{ST}), and analogous measures based on 17 microsatellite DNA loci (F_{ST}). The horizontal bars indicate 95% confidence intervals for the F_{ST} and Q_{ST} estimates. Combined effects of natural selection and random drift determine the Q_{ST} estimates, whereas the F_{ST} estimates are determined by drift processes¹⁶. The dashed vertical bars indicate the effect of drift (left-hand side) and selection (right-hand side) on the quantitative traits that have been influenced by divergent selection.

Table 2 Genetic diversity and demography of populations

	<i>n</i>	<i>A</i> (s.d.)	<i>H_o</i> (s.d.)	<i>N_e</i> (95% CI)	<i>N₀/N₁</i> (95% CI)
Les	52	1.9 (±1.1)	0.17 (±0.23)	88.9*	0.003 (0.0003–0.03)
Ht	48	1.5 (±0.9)	0.14 (±0.25)	24.2 (12.1–42.2)	0.006 (0.0003–0.05)
ØM	49	1.7 (±0.9)	0.17 (±0.24)	85.0 (36.0–170.5)	0.0006 (0.00008–0.01)
Aur	28	1.6 (±0.9)	0.14 (±0.20)	55.4 (24.8–110.0)	0.001 (0.0003–0.005)

A is the mean allele number within populations. *H_o* is the mean observed heterozygosity within populations.

*Indicates the lowest sampled point of effective population size (N_e) out of 100,000 MCMC replicates. The maximum likelihood estimate of N_e could not be obtained, most probably owing to near-zero level of drift in the Les population (M. A. Beaumont, personal communication). N_0/N_1 is the ratio of current population size (N_0) and historical population size (N_1 ; Supplementary Information)²⁹.

models emphasizing the importance of drift in small or bottlenecked populations^{2,6,8,9}. Our results are especially pertinent to this discussion (at the phenotypic level of adaptive differentiation) because: (1) the effect of drift is also estimated; and (2) the results are considered in the light of population demography. That selection often clearly outweighed drift is striking because the *T. thymallus* populations originated from a small number of founders, and were thus predisposed to the influence of drift^{2,6,8,9}. In line with the known demographic history, we also found that the populations exhibit very small microsatellite-based N_e estimates and show signatures of severe bottlenecks (Table 2). The low effective population sizes and bottlenecks were exemplified by the Hårrtjonn grayling which had an average of only 1.5 (s.d., ± 0.9) alleles per locus (A), a harmonic mean N_e of 24.2 (95% confidence interval (CI) = 12.1–42.2) individuals since the founding of the population and a current population size of 0.6% (95% CI = 0.03–5%) of its historical size (Table 2). As a comparison, *T. thymallus* populations are generally much more diverse, exhibiting an average of 4.0 (s.d., ± 0.8) alleles per locus, as revealed using the same set of microsatellites²². Although the estimated effective sizes are low for natural salmonid populations, they seem to be realistic because of the small ecological niches characterizing the Norwegian mountain lakes investigated. For instance, the total area of Hårrtjonn is only 0.35 km² and the lake has a single 10 m² *T. thymallus* spawning ground, thus providing a suitable habitat for only a very small reproducing population.

Evolutionary models commonly predict notable divergence owing to random drift in effectively small populations^{2,6,8,9}. In this study, microsatellites indeed revealed highly significant ($P < 0.001$) genetic differentiation between all of the populations and pairwise F_{ST} estimates from 0.05 (95% CI = 0.01–0.09) to 0.21 (95% CI = 0.10–0.34; Fig. 2; Supplementary Information). Because divergence was evident at multiple loci (Supplementary Information), it is highly unlikely that the F_{ST} estimates were solely due to physical linkages to genes under selection. We therefore conclude that, although the populations evolved dominantly due to selection, drift had an effect on the quantitative trait divergence.

Are such high levels of drift theoretically feasible with constant N_e , or must founder events, fluctuating population sizes, or even selection be invoked to explain the F_{ST} estimates? We used computer simulated genotypes to address this question. When applying the known coalescence times and the microsatellite-based N_e estimates, it was evident that the empirically obtained F_{ST} estimates were theoretically expected, even though the populations shared common ancestors as few as 11.8 generations ago (Supplementary Information). In other words, the effective population sizes were sufficiently low to generate the observed levels of neutral divergence, even without variation in the number of breeders (or selection) throughout the independent evolutionary histories of the grayling populations (Supplementary Information).

Thus we have shown that selection was the dominant diversifying mechanism during the early stages of grayling life-history evolution. The power of natural selection is striking because the populations originate from a small number of founders, exhibit small effective sizes and show genetic imprints of bottlenecks, which are all conditions expected to constrain selection and favour the influence of drift^{2,6,8,9}. This study demonstrates a very clear case of fisherian evolution across a contemporary timescale. □

Methods

Population differences based on quantitative traits

The seven early life-history traits studied were: (1) incubation time (number of days from fertilization to hatching); (2) fork length (length from the tip of the snout to the tip of the median rays of the tail in mm) at hatching; (3) yolk-sac volume in mm³ (estimated as $L_{YS}H_{YS}^2(\pi/6)$, where L_{YS} and H_{YS} are the length and height of yolk-sac in mm, respectively); (4) fork length at swim-up (L_{SU}); measured when yolk-sac is resorbed and an individual begins exogenous feeding); (5) fork length at termination (L_{term}), measured at 180 degree-days (ΔD = mean temperature in °C $\times$ number of days); (6) specific growth

rate (mm/ ΔD) $G = 100[(\ln L_{term} - \ln L_{SU})/\Delta D]$; and (7) survival (%) during 180 degree-days of exogenous feeding.

The level of quantitative genetic differentiation (Q_{ST}) was estimated as follows: $Q_{ST} = \sigma_{GB}^2/(\sigma_{GB}^2 + 2\sigma_{GW}^2)$, where σ_{GB}^2 and σ_{GW}^2 are additive genetic variances between and within populations, respectively¹⁶. Restricted maximum likelihood (REML) estimates of σ_{GB}^2 and σ_{GW}^2 were obtained with programs S-PLUS 6.0 and SAS 8.0 from a common-garden experiment using a hierarchical half-sib design, where each male was mated with four unique females, yielding up to 28 families per population. The experiment involved three temperatures, mimicking the natural nursery conditions for the three populations involved²¹. Consequently, σ_{GB}^2 and σ_{GW}^2 were extracted from a nested mixed-model analysis of variance (ANOVA) as recommended²³. Population, male and female were treated as random effects, and temperature, temperature $\times$ population and temperature $\times$ male as fixed effects. In order to avoid non-genetic maternal sources of variance, the σ_{GW}^2 estimates were derived from the male variance component only. The σ_{GW}^2 estimates were multiplied with four to take into account the half-sib design.

A two-model ANOVA approach¹⁸ was used to test whether maternal effects (egg size) significantly contributed to among-population divergence. No evidence for among-population maternal effects was obtained ($F_{2,4} = 0.002$ – 0.924 , $P > 0.5$). Non-parametric bootstrapping at the family level was used to assess 95% confidence intervals for the Q_{ST} estimates. The strength of directional selection was estimated using the standardized selection differentials (s' ; in s.d. units per generation) obtained from the ratio of the trait- and environment-specific haldane estimates¹⁸ and the corresponding narrow-sense heritability estimates (h^2 ; ref. 24). The narrow-sense heritability was estimated from $4\sigma_A^2/\sigma_P^2$, where σ_A^2 is the population and temperature-specific additive genetic variance (male component of variance) and σ_P^2 is the corresponding total phenotypic variance derived from nested ANOVAs (REML method²¹).

Microsatellite analyses

Spawning grayling were sampled between the years 1995 and 2000. The individuals were genotyped with 17 microsatellite loci applying previously described protocols²⁵. All of the loci are highly variable in *T. thymallus*²². Deviations from Hardy–Weinberg equilibrium and linkage equilibrium were tested for each population and locus using the exact probability tests implemented in computer program GENEPOP 3.2a²⁶. No deviations were observed, even if the P -values were not adjusted for multiple significance tests. Estimates of interpopulation differentiation (F_{ST})¹⁷ and their 95% confidence intervals were obtained using the variance component approach (providing the F_{ST} estimator theta²⁶) and bootstrapping of loci with program FSTAT 2.9.3.1²⁷.

Harmonic mean effective sizes of the populations (N_e) were estimated using a coalescent theory maximum likelihood approach that uses a Metropolis–Hastings algorithm to estimate the range of N_e values consistent with the data (Supplementary Information)²⁸. Past variation in the sizes of the populations was estimated using a Markov Chain Monte Carlo (MCMC) simulation method²⁹. The approach assesses the most probable demographic histories of populations using the estimator $r = N_0/N_1$, where N_0 and N_1 are present and past population sizes, respectively (Supplementary Information)²⁹. Computer simulations were used to assess theoretically expected F_{ST} values with the known coalescence times, and the maximum likelihood estimates and the 95% confidence intervals of effective population sizes (Supplementary Information)³⁰.

Causes of phenotypic microevolution

The null-hypothesis of evolution by random drift was tested as follows: $F = (N_e\sigma_{GB}^2)/(h^2\sigma_{GW}^2)$, where N_e is the maximum likelihood estimate of effective population size and h^2 the narrow-sense heritability of a trait in a given population and environment^{6,7}. The numerator and denominator degrees of freedom are the number of populations compared minus one; and infinity, respectively⁷. We applied the mean of the maximum likelihood N_e estimates of Ht and Aur when testing for neutrality of their divergence. A maximum likelihood estimate of N_e could not be obtained for the Les population, and the N_e of Ht or Aur was applied when testing for neutrality of the evolution of Les–Ht and Les–Aur, respectively. Because the instability of the Metropolis–Hastings sampler was most probably due to a near-zero level of drift in the Les population (M. A. Beaumont, personal communication), the neutrality test results involving Les are conservative. The neutrality test equation⁷ was also applied to estimate the extent of N_e required to produce an F -value leading to rejection of the null-hypothesis at the $P \leq 0.05$ significance level ($F_{1,\infty} = 3.84$). Population differences based on analogous measures obtained from quantitative traits (Q_{ST}) and microsatellites (F_{ST}) were compared¹⁶. Neutral divergence of quantitative traits will result in equal estimates of Q_{ST} and F_{ST} . The magnitude of differences between Q_{ST} and F_{ST} can be used to infer the relative contribution of selection and drift on phenotypic divergence¹⁶.

Received 13 June; accepted 11 July 2002; doi:10.1038/nature01029.

1. Fisher, R. A. *The Genetical Theory of Natural Selection* (Oxford Univ. Press, Oxford, 1930).
2. Wright, S. Evolution in mendelian populations. *Genetics* **16**, 97–159 (1931).
3. Kimura, M. *The Neutral Theory of Molecular Evolution* (Cambridge Univ. Press, Cambridge, 1983).
4. Gillespie, J. H. *The Causes of Molecular Evolution* (Oxford Univ. Press, Oxford, 1991).
5. Schluter, D. *The Ecology of Adaptive Radiation* (Oxford Univ. Press, Oxford, 2000).
6. Lande, R. Natural selection and random genetic drift in phenotypic evolution. *Evolution* **30**, 314–334 (1976).
7. Lande, R. Statistical tests for natural selection on quantitative characters. *Evolution* **31**, 442–444 (1977).
8. Lynch, M. The rate of morphological evolution in mammals from the standpoint of the neutral expectation. *Am. Nat.* **136**, 727–741 (1990).
9. Carson, H. L. & Templeton, A. R. Genetic revolutions in relation to speciation phenomena: the founding of new populations. *Annu. Rev. Ecol. Syst.* **15**, 97–131 (1984).
10. Losos, J. B., Warheit, K. I. & Schoener, T. W. Adaptive differentiation following experimental island colonization in *Anolis* lizards. *Nature* **387**, 70–73 (1997).

11. Reznick, D. N., Shaw, F. H., Rodd, F. H. & Shaw, R. G. Evaluation of the rate of evolution in natural populations of guppies (*Poecilia reticulata*). *Science* **275**, 1934–1936 (1997).
12. Majerus, M. E. N. *Melanism: Evolution in Action* (Oxford Univ. Press, Oxford, 1998).
13. Hendry, A. P. Adaptive divergence and the evolution of reproductive isolation in the wild: an empirical demonstration using introduced sockeye salmon. *Genetica* **112**, 515–534 (2001).
14. Grant, P. R. & Grant, B. R. Unpredictable evolution in a 30-year study of Darwin's finches. *Science* **296**, 707–711 (2002).
15. Baker, A. J. Genetic and morphometric divergence in ancestral European and descendent New Zealand populations of chaffinches (*Fringilla coelebs*). *Evolution* **46**, 1784–1800 (1992).
16. Merilä, J. & Crnokrak, P. Comparison of genetic differentiation at marker loci and quantitative traits. *J. Evol. Biol.* **14**, 892–903 (2001).
17. Wright, S. The genetical structure of populations. *Ann. Eugen.* **15**, 323–354 (1951).
18. Haugen, T. O. & Vollestad, L. A. A century of life-history evolution in grayling. *Genetica* **112**, 475–491 (2001).
19. Lynch, M. & Walsh, B. *Genetics and Analysis of Quantitative Traits* 328–352 (Sinauer, Sunderland, 1998).
20. Kingsolver, J. G. *et al.* The strength of phenotypic selection in natural populations. *Am. Nat.* **157**, 245–261 (2001).
21. Haugen, T. O. & Vollestad, L. A. Population differences in early life-history traits in grayling. *J. Evol. Biol.* **13**, 897–905 (2000).
22. Koskinen, M. T. *et al.* Microsatellite data resolve phylogeographic patterns in European grayling, *Thymallus thymallus*, Salmonidae. *Heredity* **88**, 391–401 (2002).
23. Fry, J. D. The mixed-model analysis of variance applied to quantitative genetics: biological meaning of the parameters. *Evolution* **46**, 540–550 (1992).
24. Hendry, A. P. & Kinnison, M. T. The pace of modern life: measuring rates of contemporary microevolution. *Evolution* **53**, 1637–1653 (1999).
25. Koskinen, M. T. & Primmer, C. R. High throughput analysis of 17 microsatellite loci in grayling (*Thymallus* spp. Salmonidae). *Conserv. Genet.* **2**, 173–177 (2001).
26. Raymond, M. & Rousset, F. GENEPOP (version 1.2): population genetics software for exact tests and ecumenicism. *J. Hered.* **86**, 248–249.
27. Goudet, J. FSTAT (vers. 1.2): a computer program to calculate F-statistics. *J. Hered.* **86**, 485–486 (1995).
28. O'Ryan, C. *et al.* Microsatellite analysis of genetic diversity in fragmented South African buffalo populations. *Anim. Conserv.* **1**, 85–94 (1998).
29. Beaumont, M. A. Detecting population expansion and decline using microsatellites. *Genetics* **153**, 2013–2029 (1999).
30. Balloux, F. EASYPOP (version 1.7): a computer program for population genetics simulations. *J. Hered.* **92**, 301–302 (2001).

Supplementary Information accompanies the paper on Nature's website
(<http://www.nature.com/nature>).

Acknowledgements We thank A. Hendry for his thorough reviews. We also thank M. Beaumont, P. Bredbacka, P. Crnokrak, D. Johnson, M. Lascoux, J. Merilä, R. O'Hara, W. Provine, N. Smith and A. Vollestad for comments. The work was supported by grants from the Biological Interactions Graduate School, University of Helsinki and the Finnish Academy.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to M.T.K.
(e-mail: mtkoskin@cc.helsinki.fi).

Paternal inheritance of a female moth's mating preference

Vikram K. Iyengar, H. Kern Reeve & Thomas Eisner

Department of Neurobiology and Behavior, Cornell University, Ithaca, New York 14850, USA

Females of the arctiid moth *Utetheisa ornatrix* mate preferentially with larger males, receiving both direct phenotypic and indirect genetic benefits¹. Here we demonstrate that the female's mating preference is inherited through the father rather than the mother, indicating that the preference gene or genes lie mostly or exclusively on the Z sex chromosome, which is strictly paternally inherited by daughters. Furthermore, we show that the preferred male trait and the female preference for that trait are correlated, as females with larger fathers have a stronger preference for larger males. These findings are predicted by the protected invasion theory^{2,3}, which asserts that male homogametic sex chromosome systems (ZZ/ZW) found in lepidopterans and

birds promote the evolution of exaggerated male traits through sexual selection. Specifically, the theory predicts that, because female preference alleles arising on the Z chromosome are transmitted to all sons that have the father's attractive trait rather than to only a fraction of the sons, such alleles will experience stronger positive selection and be less vulnerable to chance loss than would autosomal alleles.

The benefits accrued by female *Utetheisa* as a result of mating preferentially with larger males have been characterized. The phenotypic benefits take the form of nutrient and pyrrolizidine alkaloid transmitted seminally by the male to the female in quantities proportional to his size^{4–6}. The nutrient enables the female to increase her egg production⁷, and the alkaloid bestows chemical protection upon herself⁸ and her eggs^{1,5}. The genetic benefits, realized as a consequence of the heritability of body size⁹, ensure that females, by choosing larger males, have larger sons which are themselves more likely to be favoured in courtship, and larger daughters bound to be more fecund¹. The cumulative effect of these benefits is substantial: a female given a choice between males differing by 10% in body mass will have an estimated 25% more grandchildren by mating with the larger male. The strong selection for large males in *Utetheisa* may account for why in this species, contrary to the norm for Lepidoptera¹⁰, males are larger than females¹¹.

Sexual selection models generally assume heritable variation in both the trait and the mating preference, and a genetic correlation between trait and preference^{2,13}. In *Utetheisa*, the male trait (body size) is inherited from both parents⁹, but whether the female preference itself is genetically variable and a correlate of the male trait was unknown.

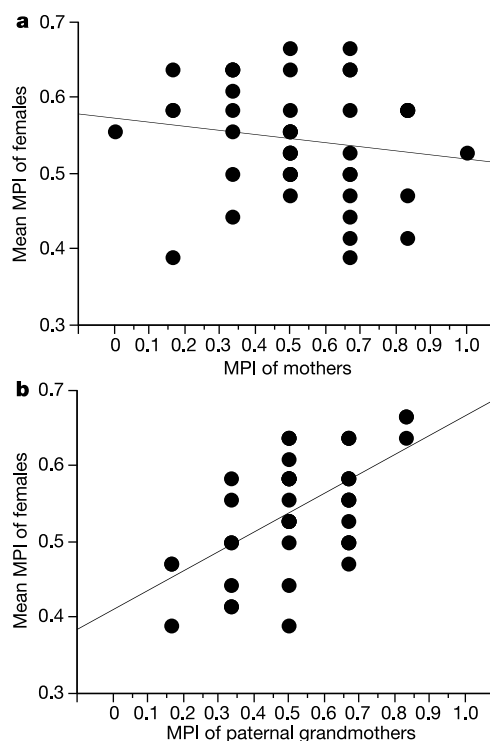


Figure 1 Mean mating preference index (MPI) of females (six full sisters) plotted as a function of the mating preference index of their mother and paternal grandmother ($n = 44$). **a**, The MPI values of females and their mothers are not correlated ($r^2 = 0.025$, $P = 0.305$, $y = 0.574 - 0.054x$), indicating that the mating preference is not inherited via the mother. **b**, The MPI values of females and their paternal grandmothers are positively correlated ($r^2 = 0.333$, $P < 0.0001$, $y = 0.41 + 0.257x$), indicating that the mating preference is paternally inherited by daughters [$h^2 = 0.513 \pm 0.112$ (mean $\pm$ s.e.)].

First, we determined whether the genetically variable components of the female mating preference are inherited through the mother, the father, or both parents. To assess inheritance through the father, who does not express the female mating preference, we focused on the transmission of the mating preference from the female's paternal grandmother. We examined inheritance patterns by calculating mating preference indices (MPIs) for three generations of females ($n = 44$ families), as follows. We subjected each female to six sequential two-male choice tests, scoring the outcome of each test as either 0 (if the smaller male mated) or 1 (if the larger male mated). The resulting six values were then averaged to determine each female's MPI score. To estimate heritability of the mating preference—that is, the proportion of phenotypic variance attributable to additive genetic effects—we regressed female MPIs on paternal grandmother MPIs and mother MPIs to determine whether mating preferences are inherited from the paternal grandmother (an indication of inheritance through the father) and/or the mother.

The MPIs of females did not correlate with those of their mothers, indicating that mating preferences are not inherited from the mother (Fig. 1a). However, the MPIs of females and their paternal grandmothers were significantly positively correlated, and the regression slope was significantly greater than zero (Fig. 1b). Moreover, the regression slopes of female–mother and female–paternal-grandmother MPI values were significantly different (analysis of covariance (ANCOVA), $F_{1,84} = 15.832$, $P < 0.0001$). These results indicate that variable female mating preference genes are inherited by daughters primarily from the father. Because the preference is sex-linked and females inherit one-half of their sex chromosomes from their paternal grandparents (see Fig. 2), heritability from the paternal grandmother was taken to be two rather than four times the regression slope¹⁴.

Second, we examined female MPI values relative to their father's body size to determine whether there was a genetic correlation between the female mating preference and the male trait. We found a significant positive correlation between female MPIs and their father's body size (Fig. 3), that is, daughters of larger fathers have a stronger preference for larger males. In contrast, there was no correlation between a female's MPI and her mother's or paternal grandmother's body size (female–mother: $r^2 = 0.008$, $n = 44$, $P = 0.56$; $y = 0.507 + 0.0004x$; female–paternal grandmother: $r^2 = 0.023$, $n = 44$, $P = 0.33$; $y = 0.593 - 0.001x$). Furthermore, a female's mating preference was not related to her own body size ($r^2 = 0.043$, $n = 44$, $P = 0.18$; $y = 0.422 + 0.001x$), as larger females were not choosier than smaller females. The positive genetic correlation between the female preference and the male trait in *Utetheisa* is required by the two major alternative models of indirect

sexual selection (fisherian and 'good genes')¹³.

Recent studies of mating preferences focused on the inheritance and sex-linkage of male traits and left unanswered the question of the chromosomal location of the gene(s) that encode the mating preference^{15–19}. *Utetheisa*, like all lepidopterans, has a 'reversed' genetic architecture (ZZ/ZW) where males are homogametic (ZZ) and females heterogametic (ZW). This sex determination system, coupled with the observed inheritance patterns for the female preference genes, that is, the daughter's receipt of such genes from the father only, permits assignment of the preference genes of *Utetheisa* to the Z chromosome. There is no evidence for the autosomal location of preference genes, as indicated by the lack of heritability between mother and daughters. In contrast, the male trait (body size) is inherited from both parents equally⁹, indicating primarily autosomal genetic determination. Given that archiids have, on average, 31 pairs of chromosomes²⁰ and that sex chromosomes in Lepidoptera are usually comparable in size to autosomes²¹, the probability of all mating preference genes being on the Z chromosome by chance is small (about $(1/62)^n$, where n is the number of genes). Thus, we conclude that the occurrence of the female mating preference gene(s) on the Z chromosome, instead of autosomes, is the consequence of a nonrandom, underlying evolutionary process.

The strong evidence for Z-linked, and not autosomal, female preference gene(s) supports the protected invasion theory, which predicts that mutant female preference genes arising on the Z chromosome will be especially protected from random loss when rare (H.K.R. and D.W. Pfennig, unpublished work). This protection arises from the fact that females transmit a Z-linked female preference gene to all their sons, at least half of which also receive the male trait that is the target of the preference. Sons with the male trait have enhanced mating success, and the enhanced mating success of the sons who also possess the female preference gene will cause that gene to rise in frequency in the next generation, reducing the probability of its random loss from the population. In contrast, no sons are guaranteed possession of both the female preference gene and the male trait gene when genes are X-linked, and only a quarter of sons have such assurance when genes are autosomal (H.K.R. and D.W. Pfennig, unpublished work). Thus, the protected invasion theory predicts that species with Z-linked genetic systems, such as lepidopterans and birds, will be especially prone to evolve exaggerated male ornaments through sexual selection (H.K.R. and D.W. Pfennig, unpublished work). Z-linked mating preferences may in fact play a role in speciation itself, as female progeny of interspecific crosses between two species of sexually dimorphic *Colias* butterflies have been shown to prefer males of the paternal species²². Thus, the Z-linkage of female mating

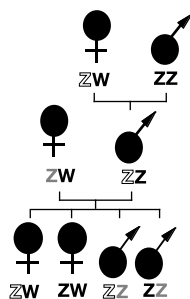


Figure 2 Inheritance of a Z-linked mating preference. The white Z represents the sex chromosome bearing the preference gene(s), shown here to be introduced into the lineage by the paternal grandmother. The grey Z is that possessed by the mother but not transmitted to daughters. Therefore, granddaughters can only acquire Z-linked gene(s) by way of their father.

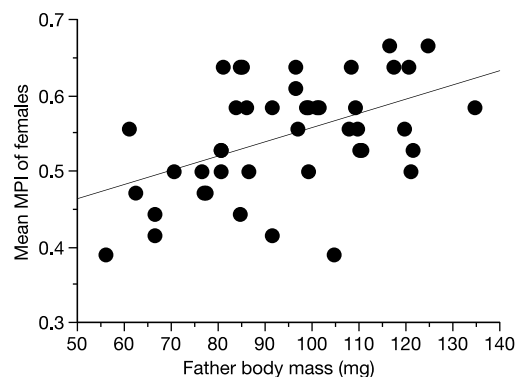


Figure 3 Correlation between mean mating preference index (MPI) of females (six full sisters) and the body mass of their father ($n = 44$). Female MPI values and their father's body mass were positively correlated ($r^2 = 0.232$, $P < 0.001$).

preferences is potentially a general phenomenon in the Lepidoptera. What we show here for *Utetheisa*, and attribute to genetic architecture, might in fact apply beyond lepidopterans and birds to any group that exhibits male homogamy. The protected invasion theory, therefore, might be invoked broadly to account for taxonomic variation in the degree to which sexual dimorphism is manifested in nature. □

Methods

Utetheisa rearing and mating

Larvae were reared as previously described²³ on a pinto-bean-based diet supplemented with seeds of *Crotalaria spectabilis*, a natural food plant of *Utetheisa*. For any set of progeny, two groups of 8–10 larvae each were raised under identical conditions to adulthood. As in earlier studies²⁴, all matings involved presenting females with a choice of two males for 24 h in humidified cylindrical containers (0.35 l). For the present purposes, the males were related neither to the female nor to each other. Males and females were of known body mass at the time of mating, and the males were all 3-day-old virgins. Events in the mating enclosures were monitored on an ongoing basis for the first hour (to ensure that both males actively courted) and then at intervals of 6 h (to check on the occurrence of mating). A record was kept of the male that mated (the males were wing-marked for recognition purposes).

The experimental protocol required that individual females be presented daily with an unfamiliar pair of males until they mated a total of six times. The two males in the pair initially offered were chosen such that, in a randomly selected half of cases, they differed by 5% in body mass, and by 10% in the remaining cases. Subsequent presentations to any one female were dependent on whether the female mated. If she did not mate, she was next presented males showing the same mass difference; if she did, her next choice was between males of the alternative mass difference.

Mating Preference Index

For experimental purposes, we raised three generations of 44 families (derived from wild *Utetheisa* females caught in Highlands County, Florida, USA). Within each family, mating preferences were assessed for six full sisters, their mother and their paternal grandmother. The mating preference index (MPI) was calculated for each female as the average of her six mating choices. Individual choices were scored as 0 if the female favoured the smaller male, and as 1 if she chose the larger male. MPI values therefore fell within the range of 0 to 1.

Statistical analyses

To estimate heritability (in the narrow sense)¹⁴, we regressed the average MPI value of six full sisters on the MPIs of their mother and paternal grandmother. A significant positive correlation was evident in the female–paternal-grandmother regression only, indicating that the mating preference is sex-linked rather than autosomal. For calculation of heritability, therefore, the slope of the regression was multiplied by two rather than four because females share half of their sex chromosomes, and only a quarter of their autosomes, with their paternal grandparents¹⁴. The data were not transformed because they met the assumptions of normality (normal probability plot) and linearity (Lowess test). Furthermore, there was no need to adjust regression coefficients or standard errors for unequal variances because the variances were the same for mother and paternal-grandmother MPI values (two-tailed variance ratio test, $F = 1.728$, $P = 0.08$). Heritabilities (regression slopes) were compared using ANCOVAs¹⁴.

Received 24 June; accepted 11 July 2002; doi:10.1038/nature01027.

1. Iyengar, V. K. & Eisner, T. Female choice increases offspring fitness in an arctiid moth (*Utetheisa ornatrix*). *Proc. Natl Acad. Sci. USA* **96**, 15013–15016 (1999).
2. Reeve, H. K. Haplodiploidy, eusociality and absence of male parental and alloparental care in Hymenoptera: a unifying genetic hypothesis distinct from kin selection theory. *Phil. Trans. R. Soc. Lond. B* **342**, 335–352 (1993).
3. Reeve, H. K. & Shellman-Reeve, J. S. The general protected invasion theory: Sex biases in parental and alloparental care. *Evol. Ecol.* **11**, 357–370 (1997).
4. Conner, W. E., Roach, B., Benedict, E., Meinwald, J. & Eisner, T. Courtship pheromone production and body size as correlates of larval diet in males of the arctiid moth, *Utetheisa ornatrix*. *J. Chem. Ecol.* **16**, 543–552 (1990).
5. Dussourd, D. E., Harvis, C., Resch, J., Meinwald, J. & Eisner, T. Pheromonal advertisement of a nuptial gift by a male moth (*Utetheisa ornatrix*). *Proc. Natl Acad. Sci. USA* **88**, 9224–9227 (1991).
6. LaMunyon, C. W. & Eisner, T. Spermatophore size as determinant of paternity in an arctiid moth (*Utetheisa ornatrix*). *Proc. Natl Acad. Sci. USA* **91**, 7081–7084 (1994).
7. LaMunyon, C. W. Increased fecundity, as a function of multiple mating, in an arctiid moth, *Utetheisa ornatrix*. *Ecol. Entomol.* **22**, 69–73 (1997).
8. González, A., Rossini, C., Eisner, M. & Eisner, T. Sexually transmitted chemical defense in a moth (*Utetheisa ornatrix*). *Proc. Natl Acad. Sci. USA* **96**, 5570–5574 (1999).
9. Iyengar, V. K. & Eisner, T. Heritability of body mass, a sexually selected trait, in an arctiid moth (*Utetheisa ornatrix*). *Proc. Natl Acad. Sci. USA* **96**, 9169–9171 (1999).
10. Opler, P. A. & Krizek, G. O. *Butterflies East of the Great Plains* (John Hopkins Univ., Baltimore, 1984).
11. LaMunyon, C. W. & Eisner, T. Postcopulatory sexual selection in an arctiid moth (*Utetheisa ornatrix*). *Proc. Natl Acad. Sci. USA* **90**, 4689–4692 (1993).
12. Lande, R. Models of speciation by sexual selection on polygenic traits. *Proc. Natl Acad. Sci. USA* **78**, 3721–3725 (1981).
13. Andersson, M. *Sexual Selection* (Princeton Univ. Press, Princeton, 1994).
14. Lynch, M. & Walsh, B. *Genetics and Analysis of Quantitative Traits*, 4th edn (Sinauer, Sunderland, 1998).

15. Houde, A. E. Sex-linked heritability of a sexually selected character in a natural population of *Poecilia reticulata* (Pisces: Poeciliidae) (guppies). *Heredity* **69**, 229–235 (1992).
16. Wilkinson, G. S., Kahler, H. & Baker, R. H. Evolution of female mating preferences in stalk-eyed flies. *Behav. Ecol.* **9**, 525–533 (1998).
17. Brooks, R. & Endler, J. A. Female guppies agree to differ: phenotypic and genetic variation in mate-choice behaviour and the consequences for sexual selection. *Evolution* **55**, 1644–1655 (2001).
18. Brooks, R. & Endler, J. A. Direct and indirect sexual selection and quantitative genetics of male traits in guppies (*Poecilia reticulata*). *Evolution* **55**, 1002–1015 (2001).
19. Wolfenbarger, L. L. & Wilkinson, G. S. Sex-linked expression of a sexually selected trait in the stalk-eyed fly, *Cryptodiopsis dalmanni*. *Evolution* **55**, 103–110 (2001).
20. Sharma, V. L. Chromosome studies on two species of moths. *Bionature* **19**, 65–67 (1999).
21. Traut, W. & Frantisek, M. Sex chromosome differentiation in some species of Lepidoptera (Insecta). *Chromosome Res.* **5**, 283–291 (1997).
22. Grula, J. W. & Taylor, O. R. Jr The effect of X-chromosome inheritance on mate selection behaviour in the sulfur butterflies, *Colias eurytheme* and *C. philodice*. *Evolution* **34**, 688–695 (1980).
23. Conner, W. E., Eisner, T., Vander Meer, R. K., Guerrero, A. & Meinwald, J. Precopulatory sexual interaction in an arctiid moth (*Utetheisa ornatrix*): Role of a pheromone derived from dietary alkaloids. *Behav. Ecol. Sociobiol.* **9**, 227–235 (1981).
24. Iyengar, V. K., Rossini, C. & Eisner, T. Precopulatory assessment of male quality in an arctiid moth (*Utetheisa ornatrix*): hydroxydanaidal is the only criterion of choice. *Behav. Ecol. Sociobiol.* **49**, 283–288 (2001).

Acknowledgements We thank J. Ladau and W. E. Conner for the collection of field-caught individuals, and J. Schlesinger for technical assistance. This manuscript was improved through comments from J. Dale, S. M. Flaxman and E. A. Tibbetts. Research support was provided by the National Institutes of Health (T.E.).

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to V.K.I. (e-mail: viyengar@wooster.edu).

Detecting recent positive selection in the human genome from haplotype structure

Pardis C. Sabeti*†‡, David E. Reich*, John M. Higgins* Haninah Z. P. Levine*, Daniel J. Richter*, Stephen F. Schaffner*, Stacey B. Gabriel*, Jill V. Planko*, Nick J. Patterson*, Gavin J. McDonald*, Hans C. Ackerman‡, Sarah J. Campbell‡, David Altshuler*§, Richard Cooper||, Dominic Kwiatkowski‡, Ryk Ward† & Eric S. Lander*¶

* Whitehead Institute/MIT Center for Genome Research, Nine Cambridge Center, Cambridge, Massachusetts 02142, USA

† Institute of Biological Anthropology, University of Oxford, Oxford, OX2 6QS, UK

‡ Wellcome Trust Centre for Human Genetics, Roosevelt Drive, Oxford, OX3 7BN, UK

§ Departments of Genetics and Medicine, Harvard Medical School, Department of Molecular Biology and Diabetes Unit, Massachusetts General Hospital, Boston, Massachusetts 02114, USA

|| Department of Preventive Medicine and Epidemiology, Loyola University Medical School, Maywood, Illinois 60143, USA

¶ Department of Biology, MIT, Cambridge, Massachusetts 02139, USA

Harvard Medical School, Boston, Massachusetts 02115, USA

The ability to detect recent natural selection in the human population would have profound implications for the study of human history and for medicine. Here, we introduce a framework for detecting the genetic imprint of recent positive selection by analysing long-range haplotypes in human populations. We first identify haplotypes at a locus of interest (core haplotypes). We then assess the age of each core haplotype by the decay of its association to alleles at various distances from the locus, as measured by extended haplotype homozygosity (EHH). Core haplotypes that have unusually high EHH and a high population frequency indicate the presence of a mutation that rose to prominence in the human gene pool faster than expected under

neutral evolution. We applied this approach to investigate selection at two genes carrying common variants implicated in resistance to malaria: *G6PD*¹ and *CD40* ligand². At both loci, the core haplotypes carrying the proposed protective mutation stand out and show significant evidence of selection. More generally, the method could be used to scan the entire genome for evidence of recent positive selection.

The recent history of the human population is characterized by great environmental change and emergent selective agents³. The domestication of plants and animals at the start of the Neolithic, roughly 10,000 years ago, yielded an increase in human population density. Humans were confronted with the spread of new infectious diseases, new food sources and new cultural environments. The last 10,000 years have thus been some of the most interesting times in human biological history, and may be when many important genetic adaptations and disease resistances arose.

We sought to design a powerful approach for detecting recent selection. Our method relies on the relationship between an allele's frequency and the extent of linkage disequilibrium (LD) surrounding it. (LD often refers to association between two alleles. Here, we use it to measure the association between a single allele at one locus with multiple loci at various distances.) Under neutral evolution, new variants require a long time to reach high frequency in the population, and LD around the variants will decay substantially during this period owing to recombination^{4,5}. As a result, common alleles will typically be old and will have only short-range LD. Rare alleles may be either young or old and thus may have long- or short-range LD. The key characteristic of positive selection, however, is that it causes an unusually rapid rise in allele frequency, occurring over a short enough time that recombination does not substantially break down the haplotype on which the selected mutation occurs. A signature of positive natural selection is thus an allele having unusually long-range LD given its population frequency. The decay of LD, and therefore the relative scale of 'short'- and 'long'-range LD, is dependent on local recombination rates. A general test for selection on the basis of these principles must therefore control for local variation in recombination rates.

We developed an experimental design to detect positive selection

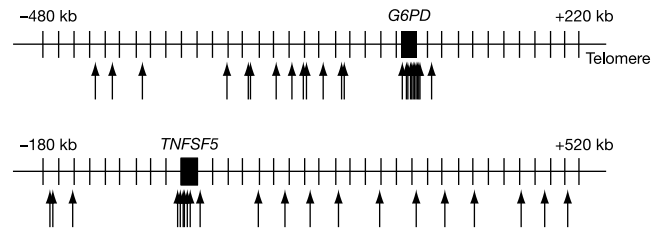


Figure 1 Experimental design of core and long-range SNPs for *G6PD* and *TNFSF5*. The core region is highlighted by a cluster of densely spaced SNPs (arrows) at the gene. Additional, widely separated flanking SNPs, used to examine the decay of LD from each core haplotype, are also shown. Markers distal to *G6PD* were within repetitive subtelomeric sequence and could not be genotyped.

at a locus using the breakdown of LD as a clock for estimating the ages of alleles. We began by genotyping a collection of single nucleotide polymorphisms (SNPs) in a small 'core region' to identify the 'core haplotypes'. We selected SNPs of sufficient density, so that recombination between them would be extremely rare and the core haplotypes could be explained in terms of a single gene genealogy (Supplementary Fig. 1). Zones of very low historical recombination were identified by looking for clusters of SNPs where Hudson's R_M was 0 and $|D'|$ was one^{6,7} (see Supplementary Fig. 1).

We then added increasingly distant SNPs to study the decay of LD from each core haplotype. To visualize this process, we generated haplotype bifurcation diagrams that branch to reflect the creation of new, extended haplotypes by historical recombination proximal and distal to the core region. We measured LD at a distance x from the core region by calculating the extended haplotype homozygosity (EHH). EHH is defined as the probability that two randomly chosen chromosomes carrying the core haplotype of interest are identical by descent (as assayed by homozygosity at all SNPs⁸) for the entire interval from the core region to the point x . EHH thus detects the transmission of an extended haplotype without recombination. Our test for positive selection involves finding a core haplotype with a combination of high frequency and high EHH, as compared with

Table 1 Core haplotype frequencies in six populations

(a) <i>G6PD</i>																		
Core haplotype	Core SNP alleles (kb)										Core haplotype frequencies in six populations							
	-10	-2	-2	-1	0	1	1	3	3	4	4	Total	Beni	Yoruba	Shona	African American	European American	Asian
1	C	G	C	G	G	A	C	C	G	C	C	0.13 (54)	0.15 (9)	0.21 (18)	0.13 (11)	0.13 (12)	0.03 (2)	0.07 (2)
2	-	-	-	-	-	-	-	-	-	-	T	0.16 (67)	0 (0)	0 (0)	0 (0)	0.07 (6)	0.57 (37)	0.80 (24)
3	-	-	-	-	-	-	-	-	-	T	-	0.25 (102)	0.23 (14)	0.24 (21)	0.45 (37)	0.19 (17)	0.17 (11)	0.07 (2)
4	-	-	-	-	-	-	-	T	-	-	-	0.01 (5)	0.02 (1)	0 (0)	0.04 (3)	0.01 (1)	0 (0)	0 (0)
5	-	-	-	-	-	-	-	T	A	-	-	0.10 (41)	0.22 (13)	0.11 (10)	0.06 (5)	0.14 (13)	0 (0)	0 (0)
6	-	-	-	-	-	G	-	-	-	-	-	0.12 (48)	0.17 (10)	0.10 (9)	0.11 (9)	0.22 (20)	0 (0)	0 (0)
7	-	-	T	A	-	G	G	-	-	-	-	0.04 (17)	0.05 (3)	0.07 (6)	0.06 (5)	0.03 (3)	0 (0)	0 (0)
8	-	A*	T	A	A	G	G	-	-	-	-	0.13 (53)	0.17 (10)	0.23 (20)	0.13 (11)	0.13 (12)	0 (0)	0 (0)
9	T	-	-	-	-	-	-	-	-	-	T	0.07 (28)	0 (0)	0.03 (3)	0.02 (2)	0.07 (6)	0.23 (15)	0.07 (2)
											N	415	60	87	83	90	65	30

(b) <i>TNFSF5</i>																		
Core haplotype	Core SNP alleles (kb)					Core haplotype frequencies in six populations												
	-6	0	1	3	4	Total	Beni	Yoruba	Shona	African American	European American	Asian						
1	T	T	T	T	G	0.03 (12)	0.06 (4)	0.03 (3)	0.02 (2)	0.04 (3)	0 (0)	0 (0)						
2	C	-	-	-	-	0.50 (200)	0.32 (20)	0.38 (33)	0.46 (38)	0.40 (32)	0.78 (49)	1.00 (28)						
3	C	-	-	C	-	0.08 (32)	0.10 (6)	0.12 (10)	0.06 (5)	0.14 (11)	0 (0)	0 (0)						
4	-	C*	-	-	-	0.25 (100)	0.38 (24)	0.38 (33)	0.26 (21)	0.27 (22)	0 (0)	0 (0)						
5	-	-	C	-	-	0.12 (47)	0.14 (9)	0.07 (6)	0.18 (15)	0.14 (11)	0.10 (6)	0 (0)						
6	-	-	C	-	A	0.02 (10)	0 (0)	0 (0)	0.01 (1)	0.01 (1)	0.13 (8)	0 (0)						
7	-	C*	C	-	A	0 (1)	0 (0)	0.01 (1)	0 (0)	0 (0)	0 (0)	0 (0)						
8	C	-	C	-	-	0 (1)	0 (0)	0 (0)	0 (0)	0.01 (1)	0 (0)	0 (0)						
					N	403	63	86	82	81	63	28						

Observed core haplotypes at *G6PD* and *TNFSF5* in six populations of African, European and Asian descent. Relative distances of core SNP alleles from the putative malaria resistance mutations are given in kb. Frequencies for haplotypes (and numbers of observations) are given for all populations. There are no apparent recombinants among the *G6PD* core haplotypes, and R_M is 0. There are 2 recombinant haplotypes among the 403 *TNFSF5* chromosomes, and R_M would also be 0 if the 2 haplotypes appearing only once were removed from the analysis⁹.

*Both proposed mutations associated with malaria resistance (*G6PD*-202A and *TNFSF5*-726C) are observed only in Africans and occur on *G6PD*-CH8 and on *TNFSF5*-CH7, respectively.

other core haplotypes at the locus. An attractive aspect of this approach is that the various core haplotypes at a locus serve as internal controls for one another, adjusting for any unevenness in the local recombination rate.

We applied our approach to two genes that have been implicated in resistance to the malaria parasite *Plasmodium falciparum*. Glucose-6-phosphate dehydrogenase (*G6PD*) is a classical example of a

gene where variants can confer malaria resistance⁹. Evidence over the past 40 years has shown that the common variant *G6PD*-202A confers partial protection against malaria, with a case-control study estimating a reduction in disease risk of about 50% (ref. 1). The CD40 ligand gene (*TNFSF5*) encodes a protein with a critical role in immune response to infectious agents. One case-control study suggested that a common variant in the promoter region,

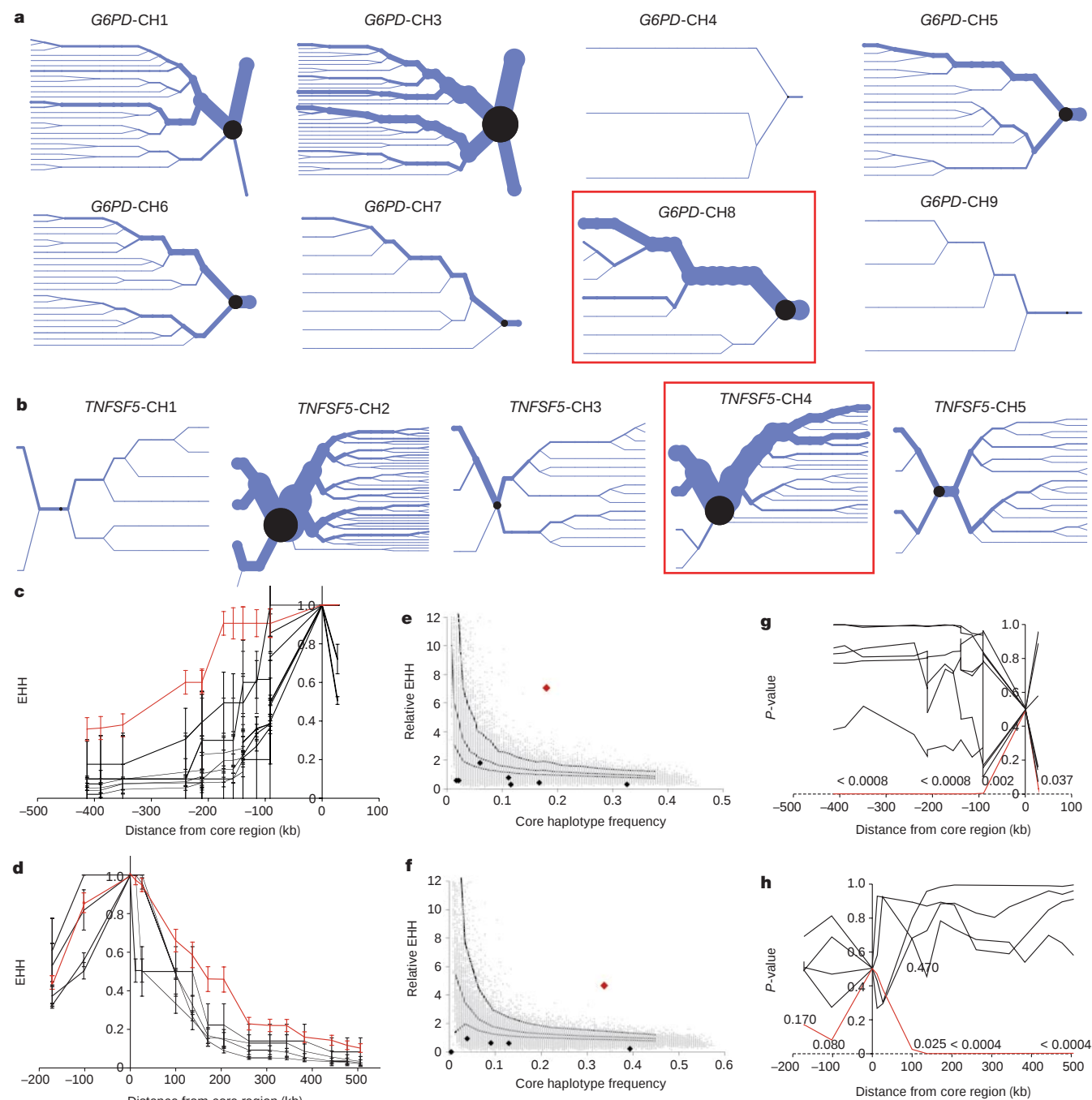


Figure 2 Core haplotype frequency and relative EHH of *G6PD* and *TNFSF5*.

a, b, Haplotype bifurcation diagrams (see Methods) for each core haplotype at *G6PD* (**a**) and *TNFSF5* (**b**) in pooled African populations demonstrate that *G6PD*-CH8 and *TNFSF5*-CH4 (boxed or labelled in red) have long-range homozygosity that is unusual given their frequency. **c, d**, The EHH at varying distances from the core region on each core haplotype at *G6PD* (**c**) and *TNFSF5* (**d**) demonstrates that *G6PD*-CH8 and *TNFSF5*-CH4 have persistent, high EHH values. **e, f**, At the most distant SNP from *G6PD* (**e**) and *TNFSF5* (**f**) core regions, the relative EHH plotted against the core haplotype frequency is presented

and compared with the distribution of simulated core haplotypes (on the basis of simulation of 5,000 data sets; represented by grey dots and given with 95th, 75th and 50th percentiles). The observed non-selected core haplotypes in our data are represented by black diamonds. **g, h**, We calculated the statistical significance of the departure of the observed data from the simulated distribution at each distance from the core. *G6PD*-CH8 (**g**) and *TNFSF5*-CH4 (**h**) demonstrate increasing deviation from a model of neutral drift at further distances from the core region in both directions.

TNFSF5-726C, is associated with a similar degree of protection against malaria².

We first studied *G6PD* (Fig. 1). We defined a core region of 15 kilobases (kb) at *G6PD* and genotyped 11 SNPs in 3 African and 2 non-African populations. The SNPs defined 9 core haplotypes (Table 1a) (denoted *G6PD*-CH1 to 9, for core haplotypes 1 to 9). The *G6PD*-202A allele, which has been associated with protection from malaria, was carried on only one core haplotype, *G6PD*-CH8. Notably, *G6PD*-CH8 is common in Africa (18%), where malaria is endemic, but is absent outside of Africa. For carrying out our test for selection, we focused on the three African populations, which

did not differ significantly with respect to core haplotype frequencies (by Fisher's exact test¹⁰) and hence were pooled for the main analysis. (Analyses were also performed separately for each population, yielding qualitatively similar results; see below.)

G6PD-CH8 demonstrates clear long-range LD (as seen by the predominance of one thick branch in the haplotype bifurcation diagram (Fig. 2a)) and has correspondingly high EHH. The EHH is 0.38 at the largest distance tested; that is, 413 kb (Fig. 2c). For each core haplotype we calculated the relative EHH; specifically, the factor by which EHH decays on the tested core haplotype compared with the decay of EHH on all other core haplotypes combined (Methods).

To test formally for selection, for each core haplotype, we compared the allele frequency to the relative EHH at various distances (Fig. 2e shows the comparison at 413 kb proximal to *G6PD*). *G6PD*-CH8 has a much higher relative EHH than other haplotypes of comparable frequency, but is this statistically significant? To obtain a sense of how unusual our observation is, we simulated haplotypes using a coalescent process (see Methods and Fig. 2e)¹¹. The deviation from the simulation results is highly significant and becomes progressively more marked with increasing distance (Fig. 2g) (*P*-values at 413 kb proximal are: constant-sized population, *P* < 0.0008; expansion, *P* < 0.0006; bottleneck, *P* < 0.0008; population structure, *P* < 0.0008; see Methods and Supplementary Table 2 for details of the demographic models we considered). The frequency and LD properties of *G6PD*-CH8 are incompatible with what is expected under a model of neutral evolution for a wide range of demographics. Furthermore, when the three African populations comprising the pooled sample were considered separately, a signal of selection was identified independently in each population (Yoruba, *P* < 0.0012; Beni, *P* < 0.0440; and Shona, *P* < 0.0030, based on simulation of a constant-sized population), demonstrating that the signal of selection is not an artefact of pooling the three population samples.

We next applied our approach to the CD40 ligand gene (Fig. 1). We defined a core region of 10 kb and genotyped 5 SNPs. The SNPs defined seven core haplotypes (Table 1b). The *TNFSF5*-726C allele, which has been associated with protection from malaria, was present on *TNFSF5*-CH4, which is common in Africa (34%), but is absent outside of Africa. *TNFSF5*-CH4 demonstrates high LD as seen in the haplotype bifurcation diagrams (Fig. 2b) and has high EHH at long distances (Fig. 2d). *TNFSF5*-CH4 is a clear outlier when compared with other haplotypes (Fig. 2f), and its frequency and LD properties are incompatible with neutral evolution under multiple demographic models (*P*-values at 506 kb distal are: constant-sized population, *P* < 0.0012; expansion, *P* < 0.0008; bottleneck, *P* < 0.0012; population structure, *P* < 0.0008; see Methods and Supplementary Table 2 for details). Again, the *P*-value is increasingly significant at further distances from the core region both proximally and distally (Fig. 2h). When each African population was analysed separately, a signal of selection was significant (Yoruba, *P* < 0.0008; Beni, *P* < 0.0023; Shona, *P* < 0.0242). These results thus provide independent evidence supporting the proposed role of CD40 ligand in malaria resistance².

We tested our conclusion of positive selection by performing a similar analysis on 17 randomly chosen control regions across the human genome in the same African populations. We only used data from each control if it was closely matched to our data in terms of the number of chromosomes studied and the homozygosity at the core haplotype and at long distances from the core (Fig. 3). *G6PD*-CH8 and *TNFSF5*-CH4 clearly stand out from the other loci, showing that the *P*-values determined by simulation are also supported by direct, empirical comparison. In measuring *P*-values for the controls where there is no prior hypothesis of selection, a Bonferroni correction for multiple-hypothesis testing was applied. Notably, one core haplotype, from the monocyte chemotactic protein 1 region, shows frequency and LD properties similar to

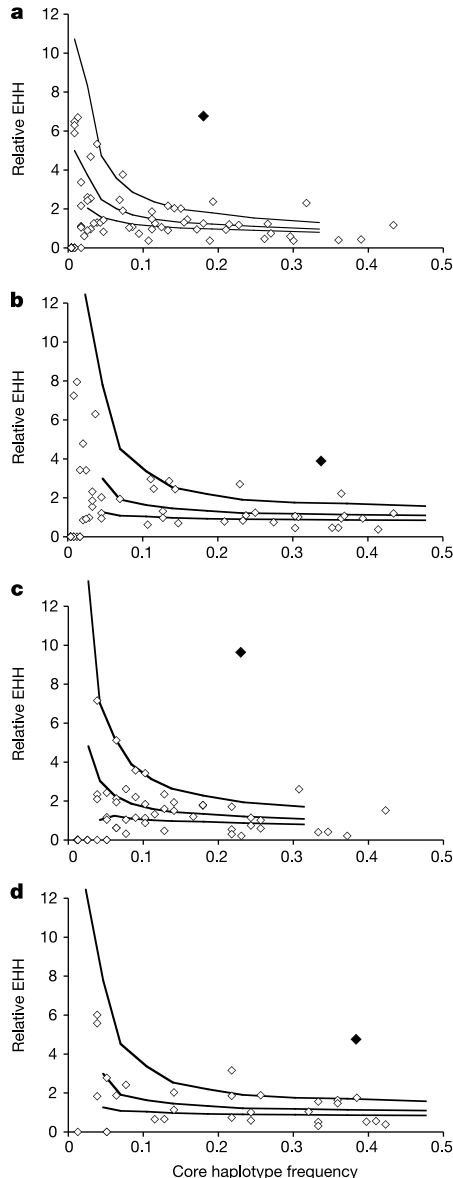


Figure 3 Control regions: core haplotype frequency against relative EHH. To provide an empirical, non-simulation-based evaluation of the signal of selection, we compared the frequency and relative EHHs for *G6PD* (a) and *TNFSF5* (b) with patterns observed in randomly chosen genes in the genome (see Methods). c, d, We performed the entire analysis again on *G6PD* (c) and *TNFSF5* (d) for a subset of 78 Yoruban haplotypes (using family trios where phase could be determined). We were able to match 30 to 87 core haplotypes (indicated by outlined diamonds) from the control regions to our data. The 95th, 75th and 50th percentiles for simulated data are also shown. *G6PD*-CH8 and *TNFSF5*-CH4 (indicated by black diamonds) clearly stand out from the pattern seen at other loci in the genome, suggesting a true signal of selection.

G6PD and *TNFSF5*, although this nominally significant result may be simply a false-positive owing to the large number of hypotheses examined.

We used a linkage-disequilibrium-based technique¹² to estimate dates of origin of the two resistance variants. The estimates were about 2,500 years for *G6PD* and about 6,500 years for *TNFSF5* (see Supplementary Information for details). The date for *G6PD* is consistent with a recent independent age estimate for *G6PD*-202A based primarily on microsatellite data¹³.

Finally, we explored whether positive selection could have been detected with traditional tests (Supplementary Table 3)¹⁴. We performed Tajima's *D*-test¹⁵, Fu and Li's *D*-test¹⁶, Fay and Wu's *H*-test¹⁷, the Ka/Ks test¹⁸, the McDonald and Kreitman test¹⁹, and the Hudson–Kreitman–Aguadé (HKA) test²⁰. None showed significant deviation from neutral evolution for either *G6PD* or *TNFSF5*, consistent with their low power to detect recent selection.

Our approach, which we refer to as the long-range haplotype (LRH) test, provides a way to detect recent positive selection by analysing haplotype structure in random individuals from a population. How far back in human history can one detect positive selection? Selective events occurring less than 400 generations ago (10,000 years assuming 25 years per generation) should leave a clear imprint at distances of over 0.25 centiMorgans (cM). The signal of such long-range LD should be distinguishable from the background extent of LD for common haplotypes in the genome²¹, which are typically tens of thousands of generations old⁴ and hence extend 0.02 cM or less. Over many tens of thousands of years, the signal of selection will become lost as recombination whittles the long-range haplotypes to the typical size of haplotype blocks in the human genome²¹.

The LRH test can be used to search for evidence of positive selection by testing each common haplotype in a gene, without prior knowledge of a specific variant or selective advantage. Once the signature of selection is found, one must then decipher its cause. The LRH test could be applied to scan the entire human genome for evidence of recent positive selection simply by applying it to each haplotype block in reference data sets from human populations, as will be collected by the Human Haplotype Map project²¹. In this fashion, it should be possible to shed light on how the human genome was shaped by recent changes in culture and environment. The LRH test should also be useful for studying selection in other organisms, including domestic animals and parasites such as the malaria parasite *P. falciparum*²². □

Methods

Human subjects

DNA samples from 252 males from Africa were used in the study: 92 Yoruba and 73 Beni from Nigeria, and 87 Shona from Zimbabwe. Additional DNA samples from 29 Yoruban trios (father–mother–child clusters) were genotyped at the 17 control regions. The Yoruba and Shona males were healthy individuals obtained as part of the International Collaborative Study of Hypertension in Blacks. The Beni samples were from civil servants in Benin City. Samples from four non-African populations and four primates were also used (Supplementary Information).

SNP genotyping

We genotyped 49 SNPs distributed around *G6PD* and 37 SNPs distributed around *TNFSF5* using mass spectrometry (Sequenom)²³. The SNPs were identified by our own re-sequencing and through previous discovery efforts^{2,24,25}. For *G6PD*, we focused on genotyping SNPs proximal to the gene, as SNPs in the repetitive subtelomeric sequence distal to *G6PD* could not be genotyped. A total of 25 SNPs around *G6PD* and 21 SNPs around *TNFSF5* were successfully genotyped and used in analysis (see Supplementary Information for details).

Haplotype bifurcation diagrams

To visualize the breakdown of LD on core haplotypes, we created bifurcation diagrams using MATLAB. The root of each diagram is a core haplotype, identified by a black circle. The diagram is bi-directional, portraying both proximal and distal LD. Moving in one direction, each marker is an opportunity for a node; the diagram either divides or not based on whether both or only one allele is present. Thus the breakdown of LD on the core haplotype background is portrayed at progressively longer distances. The thickness of the lines corresponds to the number of samples with the indicated long-distance haplotype.

Extended haplotype homozygosity and relative EHH

EHH at a distance *x* from the core region is defined as the probability that two randomly chosen chromosomes carrying a tested core haplotype are homozygous at all SNPs^s for the entire interval from the core region to the distance *x*. EHH is on a scale of 0 (no homozygosity, all extended haplotypes are different) to 1 (complete homozygosity, all extended haplotypes are the same). Relative EHH is the ratio of the EHH on the tested core haplotype compared with the EHH of the grouped set of core haplotypes at the region not including the core haplotype tested. Relative EHH is therefore on a scale of 0 to infinity.

Coalescent simulations

We used a computer program by Hudson that simulates gene history with recombination¹¹. The program was modified to generate data such as we collected. We simulated a long region of DNA (1.3 cM), with one end defined as the 'core'. We progressively added SNPs at the core until they matched our data (for the *G6PD* or *TNFSF5* core) to within $\pm 12.5\%$ in terms of the homozygosity. To mimic the SNP selection strategy used by The SNP consortium²⁵, which was the source of most of the SNPs in our study, we only included simulated SNPs in our analysis if different alleles were observed at two randomly chosen chromosomes from the sample. At longer distances, we added additional SNPs, only choosing SNPs for analysis that matched our data in terms of frequency (within a $\pm 12.5\%$ window) and also broke down EHH to the same extent as was observed in our data (within $\pm 12.5\%$).

We repeated the simulations for 5,000 data sets (each producing typically 6–8 core haplotypes) to generate many data points with which to compare our data. *P*-values were obtained by first binning the simulated data by core haplotype frequency into 30 bins of equal size, each containing about 1,000 data points. We then ranked an observed core haplotype's relative EHH compared with that of all simulated data points within the bin containing haplotypes of the same frequency—the rank determines the *P*-value. For simulations of additional demographic histories, we considered two models of expansion, an extreme bottleneck, and a highly structured population. For expansions, we simulated a population that was constant at size 10,000 until 200 or 5,000 generations ago, when it expanded suddenly by a factor of 1,000. For an extreme bottleneck, we simulated a population that was constant at size 10,000 except for a brief bottleneck (inbreeding coefficient 0.18) that occurred 800 generations ago. (An inbreeding coefficient²⁶ of 0.18 is generated by dropping the population size to 800 chromosomes for 160 generations.) For a structured population, we simulated two equal-sized populations of size *N*/2 that exchanged migrants throughout history with a probability of $1/8N$ per generation per chromosome.

Results of the simulations remained qualitatively similar when we explored additional demographics and when we varied the stringency of matching to SNP allele frequencies and to haplotype homozygosities. A comprehensive, simulation-based exploration of the LRH test will be presented elsewhere, along with explorations of the statistical power of the test and computer code for implementing the LRH test on other data sets, including those with missing or unphased data (D.E.R., manuscript in preparation).

Control regions

To obtain control data for comparison to the *G6PD* and *TNFSF5* haplotypes, we genotyped the same population samples in 17 randomly chosen autosomal genes (*ACVR2B*, *TGFB1*, *DDR1*, *GTF2H4*, *COL11A2*, *LAMB1*, *WASL*, *SLC6A12*, *KCNA1*, *ARGHD1B*, *PCI*, *PRKCB1*, *NF1*, *SCYA2*, *PA12*, *IL17R* and *HCF2*) selected previously as part of a genome-wide survey of linkage disequilibrium²⁶. For our analyses we randomly picked chromosomes to match the numbers sampled for *G6PD* and *TNFSF5*, and we only included those control regions that we could match to our data in terms of homozygosity at the core and homozygosity at long distance ($\pm 25\%$ stringency of matching). After the filtering process, we evaluated *G6PD* at 240.3 kb proximal to the gene and *TNFSF5* at 343.9 kb distal to the gene, because we could not make enough comparisons to control regions at the further distances. Seven genes matched to *G6PD* and seven genes matched to *TNFSF5*. We repeated the analysis using 78 Yoruban chromosomes for which phase information was known experimentally (because of genotyping in trios) and for which phase for the most part did not have to be inferred computationally. Six genes matched to *G6PD* and six genes matched to *TNFSF5* (see Supplementary Information for details).

Received 7 June; accepted 19 September 2002; doi:10.1038/nature01140.

Published online 9 October 2002.

1. Ruwende, C. & Hill, A. Glucose-6-phosphate dehydrogenase deficiency and malaria. *J. Mol. Med.* **76**, 581–588 (1998).
2. Sabeti, P. *et al.* CD40L association with protection from severe malaria. *Genes Immun.* **3**, 286–291 (2002).
3. Cavalli-Sforza, L. L., Menozzi, P. & Piazza, A. *The History and Geography of Human Genes* (Princeton Univ. Press, Princeton, 1994).
4. Kimura, M. *The Neutral Theory of Molecular Evolution* (Cambridge Univ. Press, Cambridge/New York, 1983).
5. Stephens, J. C. *et al.* Dating the origin of the CCR5-Delta32 AIDS-resistance allele by the coalescence of haplotypes. *Am. J. Hum. Genet.* **62**, 1507–1515 (1998).
6. Hudson, R. R. & Kaplan, N. L. Statistical properties of the number of recombination events in the history of a sample of DNA sequences. *Genetics* **111**, 147–164 (1985).
7. Lewontin, R. The interaction of selection and linkage. I. General considerations; heterotic models. *Genetics* **49**, 49–67 (1964).
8. Nei, M. *Molecular Evolutionary Genetics* Eqn. 8.4 (Columbia Univ. Press, New York, 1987).
9. Luzatto, L., Mehta, A. & Vulliamy, T. *The Metabolic & Molecular Bases of Inherited Disease* 4517–4553 (McGraw-Hill, New York, 2001).
10. Raymond, M. & Rousset, F. An exact test for population differentiation. *Evolution* **49**, 1280–1283 (1995).
11. Hudson, R. R. Properties of a neutral allele model with intragenic recombination. *Theor. Popul. Biol.* **23**, 183–201 (1983).

12. Reich, D. E. & Goldstein, D. B. *Microsatellites: Evolution and Applications* 128–138 (Oxford Univ. Press, Oxford/New York, 1999).
13. Tishkoff, S. A. *et al.* Haplotype diversity and linkage disequilibrium at human G6PD: recent origin of alleles that confer malarial resistance. *Science* **293**, 455–462 (2001).
14. Rozas, J. & Rozas, R. DnaSP version 3: an integrated program for molecular population genetics and molecular evolution analysis. *Bioinformatics* **15**, 174–175 (1999).
15. Tajima, F. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* **123**, 585–595 (1989).
16. Fu, Y. X. & Li, W. H. Statistical tests of neutrality of mutations. *Genetics* **133**, 693–709 (1993).
17. Fay, J. C. & Wu, C. I. Hitchhiking under positive Darwinian selection. *Genetics* **155**, 1405–1413 (2000).
18. Hughes, A. L. & Nei, M. Pattern of nucleotide substitution at major histocompatibility complex class I loci reveals overdominant selection. *Nature* **335**, 167–170 (1988).
19. McDonald, J. H. & Kreitman, M. Adaptive protein evolution at the Adh locus in *Drosophila*. *Nature* **351**, 652–654 (1991).
20. Hudson, R. R., Kreitman, M. & Aguade, M. A test of neutral molecular evolution based on nucleotide data. *Genetics* **116**, 153–159 (1987).
21. Gabriel, S. B. *et al.* The structure of haplotype blocks in the human genome. *Science* **23**, 2225–2229 (2002).
22. Wootton, J. C. *et al.* Genetic diversity and chloroquine selective sweeps in *Plasmodium falciparum*. *Nature* **418**, 320–323 (2002).
23. Tang, K. *et al.* Chip-based genotyping by mass spectrometry. *Proc. Natl Acad. Sci. USA* **96**, 10016–10020 (1999).
24. Vulliamy, T. J. *et al.* Linkage disequilibrium of polymorphic sites in the G6PD gene in African populations and the origin of G6PD A. *Gene Geogr.* **5**, 13–21 (1991).
25. Sachidanandam, R. *et al.* A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. *Nature* **409**, 928–933 (2001).
26. Reich, D. E. *et al.* Linkage disequilibrium in the human genome. *Nature* **411**, 199–204 (2001).

Supplementary Information accompanies the paper on Nature's website
(<http://www.nature.com/nature>).

Acknowledgements We thank B. Blumenstiel, M. DeFelice, A. Lochner, J. Moore, H. Nguyen and J. Roy for assistance in genotyping the 17 control regions. We also thank L. Gaffney, S. Radhakrishna, T. DiCesare and T. Lavery for graphics and technical support, B. Ferrell for the Beni samples, and A. Adeyemo and C. Rotimi for helping to collect the Yoruba and Shona samples. Finally, we thank M. Daly, E. Cosman, B. Gray, V. Koduri, T. Herrington and L. Peterson for comments on the manuscript. P.C.S. was supported by grants from the Rhodes Trust, the Harvard Office of Enrichment, and by a Soros Fellowship. This work was supported by grants from the National Institute of Health.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to E.S.L.
(e-mail: lander@genome.wi.mit.edu).

Voltage-sensing mechanism is conserved among ion channels gated by opposite voltages

Roope Männikkö*, Fredrik Elinder* & H. Peter Larsson†

* Department of Neuroscience, The Nobel Institute for Neurophysiology, Karolinska Institutet, SE-171 77 Stockholm, Sweden

† Neurological Sciences Institute, Oregon Health & Science University, 505 NW 185th Avenue, Beaverton, Oregon 97006

Hyperpolarization-activated cyclic-nucleotide-gated (HCN) ion channels are found in rhythmically firing cells in the brain and in the heart¹, where the cation current through HCN channels (called I_h or I_f) causes these cells to fire repeatedly². These channels are also found in non-pacing cells, where they control resting membrane properties, modulate synaptic transmission, mediate long-term potentiation, and limit extreme hyperpolarizations^{3–7}. HCN channels share sequence motifs with depolarization-activated potassium (Kv) channels, such as the fourth transmembrane segment S4^{8,9}. S4 is the main voltage sensor of Kv channels, in which transmembrane movement of S4 charges triggers the opening of the activation gate^{10–17}. Here, using cysteine accessibility methods^{10–12}, we investigate whether S4

moves in an HCN channel. We show that S4 movement is conserved between Kv and HCN channels, which indicates that S4 is also the voltage sensor in HCN channels. Our results suggest that a conserved voltage-sensing mechanism operates in the oppositely voltage-gated Kv and HCN channels, but that there are different coupling mechanisms between the voltage sensor and activation gate in the two different channels.

If S4 is the voltage sensor in HCN channels, some of its charges must move relative to the electric field across the membrane during activation. To test this hypothesis, we measured the solvent accessibility of several residues in S4 of spHCN channels⁹ (see Methods) in both open and closed channels, to look for state-dependent changes in accessibility. We introduced cysteines at eight different positions in S4 of spHCN channels (Fig. 1a, b) and assayed the accessibility of the cysteines by applying the membrane-impermeable thiol reagent MTSET ([2-(trimethylammonium)ethyl] methanethiosulphonate, bromide)^{10–12,17}.

In Fig. 1c we show a typical family of wild-type spHCN channel currents. All mutations but R341C expressed well. Figure 1d shows typical normalized conductance versus voltage curves ($G(V)$) for wild-type channels and all seven mutations. The channels with charge-neutralizing cysteine substitutions activated at more hyperpolarized potentials than did the wild-type spHCN channels. The voltage shifts caused by the cysteine substitutions were similar to the hyperpolarizing shifts caused by neutralizations of S4 charges in other HCN channels^{18,19}.

Perfusion of external and internal MTSET (100 μ M for 2 min) had very small effects on wild-type spHCN channels, mainly a <10% change in the current amplitude. However, a similar application of MTSET on the cysteine-substituted channels resulted in clear, irreversible changes in the amplitude of the current, in the activation rate of the current, or in the voltage dependence of the channels. (K329C was not clearly affected by MTSET at any potential and was, therefore, not included in the subsequent study.) The modification rate of the cysteines was measured by plotting the amplitude of the MTSET-induced change as a function

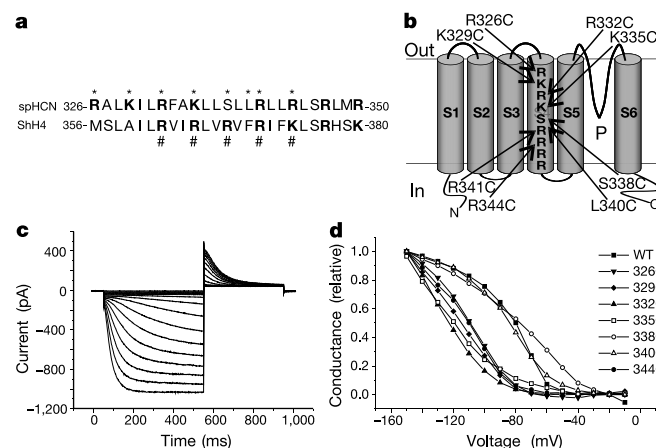


Figure 1 Basic features of spHCN channels. **a**, Sequence alignment of spHCN and Shaker channels in the S4 region. Bold indicates positive charges, asterisk indicates residues mutated in spHCN, and hash indicates the positive charges contributing to voltage sensing in Shaker^{11,13,14}. **b**, Putative transmembrane topology showing the positions of the introduced cysteines. **c**, Patch-clamp recording of currents through wild-type (WT) spHCN channels in response to voltage pulses from -10 mV to -150 mV, from a holding potential of -10 mV. Tail potential, $+50$ mV. **d**, Conductance versus voltage ($G(V)$) curves for WT and cysteine-substituted spHCN channels measured in excised patches. $G(V)$ curves measured in whole oocytes were shifted to the right by approximately 20 – 30 mV, as previously reported⁹. The conductance was calculated from tail currents, normalized as $G(V) = (I(V) - I_{\min}) / (I_{\max} - I_{\min})$.

of the cumulative exposure to MTSET (see Methods). The modification rate of the cysteines was measured for MTSET applications at both depolarized potentials (mainly closed channels) and hyperpolarized potentials (mainly open channels), to determine whether the accessibility of the S4 residues is state dependent. Changes in the accessibility of the cysteines between the open and closed states of spHCN channels were used to assess whether S4 moves across the membrane between the open and closed states of the channels.

In the sequence alignment (Fig. 1a)⁹, S338 in S4 of the spHCN channel is in the same position as R368 in S4 of the Shaker K⁺ channel. Residues at position 368 in the Shaker channel have been shown to be accessible to MTSET from the intracellular solution at hyperpolarized potentials and from the extracellular solution at depolarized potentials, suggesting that this residue moves across the membrane during gating^{11,17}. Therefore, we tested whether residues at position 338 in spHCN channels also move across the membrane by measuring the accessibility of S338C to internal and external MTSET.

We found that internal MTSET drastically altered the currents through S338C channels. MTSET removed the voltage and time dependence of the currents, but did not significantly change their maximum amplitude (Fig. 2a–c). Internal MTSET modified S338C channels >100-fold faster at hyperpolarized potentials ($V = -100$ mV) than at depolarized potentials ($V = +50$ mV) (Fig. 2d and Table 1). External MTSET also modified S338C channels, and this modification altered the currents through these channels in a similar manner to internal MTSET (Fig. 2e–g). The state dependence of the reaction rate for external MTSET was opposite to that for internally applied MTSET. External MTSET modified S338C channels 10-fold faster at depolarized potentials ($V = +50$ mV) than at hyperpolarized potentials ($V = -100$ mV) (Fig. 2h and Table 1). These results suggest that S338 moves from an externally accessible position at depolarized potentials to an internally accessible position at hyperpolarized potentials, consistent with S4 in spHCN channels moving inward when spHCN channels open. We next tested the state dependence of MTSET modification of residues around 338, to estimate the size of S4 movement in spHCN channels during gating.

The modification rate of K335C by externally applied MTSET was also found to be state dependent. The modification was >10-fold faster at depolarized potentials (+50 mV) than at hyperpolarized

potentials (−100 mV) (Table 1). K335C was not modified by internal MTSET. The modification rates of residues L340C and R344C by internal MTSET were also found to be state dependent (Table 1). Figure 3a–c shows currents before, during, and after modification of L340C with internal MTSET. The modification was 24-fold faster at hyperpolarized potentials (−100 mV) than at depolarized potentials (+50 mV) (Table 1). L340C and R344C were not modified by external MTSET. R326C and 332C were only modified by external MTSET, and the modification was not state dependent (Table 1), which confirms that the amino terminus of S4 in spHCN is always extracellular. The state-dependent accessibility of residues 335, 338, 340 and 344 is consistent with S4 in spHCN channels moving outward during depolarization and inward during hyperpolarization, which is just as S4 was found to move in the Shaker K⁺ channel^{11–17}. However, contrary to the Shaker K⁺ channel, which opens when S4 is extruded, the spHCN channel opens when S4 is retracted.

Modification of S338C, L340C and R344C by MTSET removed the voltage and time dependence of the current, but the maximum amplitude of the current remained constant (Fig. 2b and 3b). This effect was specific for these three residues, as application of MTSET did not remove the voltage and time dependence of the current through wild-type spHCN channels. It is not likely that the purely ohmic currents after MTSET modification (Fig. 3c) represent MTSET-induced unspecific leakage of the oocytes, but that these currents go through spHCN channels because the currents are still dependent on cyclic AMP and are blocked by ZD7288²⁰ (see Supplementary Information).

The purely ohmic appearance of the currents after MTSET modification suggests that the modified channels stay open permanently. One possible explanation for this is that internal MTSET modification prevents S4 from returning to its external position, thus locking the channel in the open state. Another possibility is that MTSET modification disrupts the coupling between the movement of the voltage sensor and the opening and closing of the activation gate. To evaluate these two possibilities, we tested the effect of MTSET modification on the movement of the voltage sensor.

In the Shaker K⁺ channel, the movement of S4 charges through the plane of the membrane gives rise to nonlinear, capacitive currents called gating currents^{13,14,16}. Results from our measure-

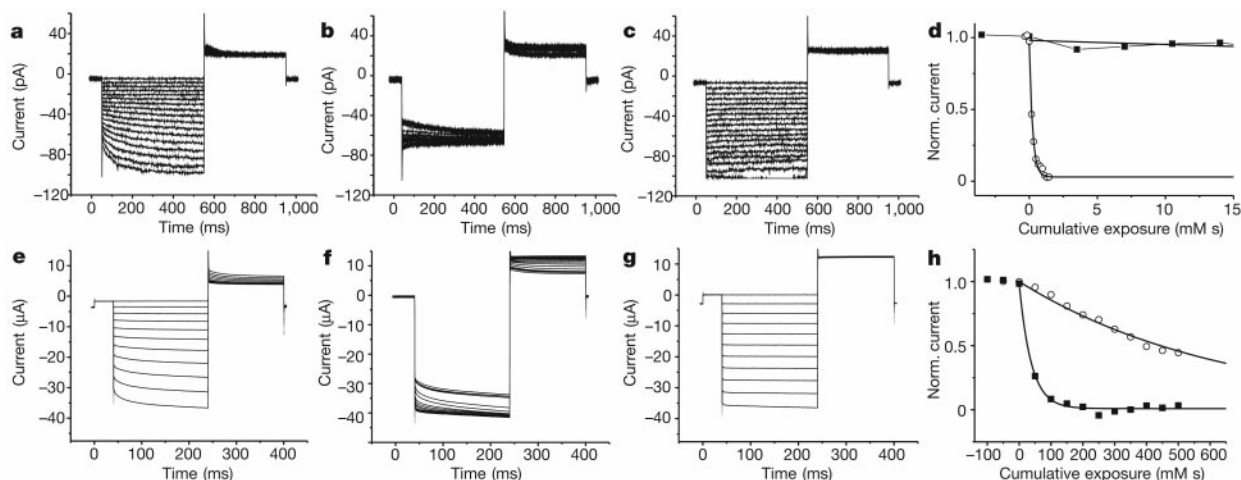


Figure 2 State-dependent accessibility of S338C. **a–c**, Currents before (**a**), during (**b**), and after (**c**) the application of 100- μ M internal MTSET. In **a** and **c**, the voltage steps ranged from −10 to −180 mV. In **b**, the test step was to −120 mV for all traces. MTSET was applied continuously. Ten seconds between each episode. Holding potential, −10 mV; tail potential, +50 mV. **d**, Modification time course for internal MTSET applied at 0 mV (squares) and at −100 mV (circles). **e–g**, Currents before (**e**), during (**f**), and after

(**g**) the application of 5 mM external MTSET. In **e** and **g**, the voltage steps ranged from 0 to −120 mV. In **f**, the test step was to −100 mV for all traces. **h**, Modification time course for external MTSET applied at +50 mV (squares) and at −100 mV (circles). The amplitude of the time-dependent current ($I_{\text{final}} - I_{\text{initial}}$) at −100 mV is plotted versus the cumulative exposure to MTSET in **d** and **h**. Thick lines are exponential fits to the data in **d** and **h**.

ments of cysteine accessibilities indicate that S4 charges in spHCN channels also move through the plane of the membrane, suggesting that the spHCN channel should also display gating currents. We set out to measure gating currents from spHCN channels expressed at high levels in *Xenopus* oocytes. To eliminate the ionic currents through the spHCN channels, we applied the specific HCN-channel blocker ZD7288²⁰ (see Methods). Nonlinear capacitive currents appeared for voltage steps more negative than -20 mV (from a holding potential of -10 mV) (Fig. 3d). When the voltage was subsequently returned to $+50$ mV, currents of similar size, but with opposite polarity, appeared. The gating charge (the integral of the gating currents) saturates at both negative and positive potentials (Fig. 3e). Uninjected oocytes did not display these nonlinear capacitive currents ($n = 3$), suggesting that they are gating currents from spHCN channels. Additional reasons why these capacitive currents are probably gating currents are presented elsewhere (see Supplementary Information).

Measuring gating currents in spHCN channels allowed us to test directly how MTSET modification of certain S4 residues locks the spHCN channel in the open state (see, for example, Fig. 2c and 3c). If MTSET modification prevents the spHCN channel from closing by preventing S4 movement, MTSET modification should also reduce the gating currents. We tested this hypothesis by measuring gating currents before and after MTSET modification of S338C channels.

MTSET applied to the extracellular solution in the two electrode voltage clamp (TEVC) configuration dramatically decreased the gating current of S338C channels ($>80\%$, $n = 7$) (Fig. 3f). The modification rate of the 338C gating current was 3.8-fold slower

than the modification rate for the 338C ionic current ($k_{\text{gating}} = 3.9 \pm 1.4 \text{ M}^{-1} \text{ s}^{-1}$ ($n = 5$) versus $k_{\text{ionic}} = 14.9 \pm 2.0 \text{ M}^{-1} \text{ s}^{-1}$ ($n = 3$), both measured at 0 mV). A similar difference in modification rates of the ionic and gating currents was found for the immobilization of S4 in Shaker K^+ channels²¹. This difference is close to what would be expected for independent movement of the four S4s in one channel²¹, under the assumption that all four S4s need to move outwards to close the channel. That the gating currents are decreased by MTSET indicates that MTSET modification of S338C reduces, or eliminates, S4 movement (in the experimental voltage range). That most of the gating current is eliminated is consistent with S4 as the main voltage sensor in the spHCN channel.

In the present study, we have found strong evidence that S4 is the voltage sensor in an HCN channel, based on the state-dependent accessibility of four S4 cysteines in spHCN channels. In the open state (at hyperpolarized potentials), a maximum of five S4 residues (333–337) are inaccessible to both extracellular and intracellular MTSET (Fig. 4a). When the spHCN channels are depolarized, a residue in this segment of S4 (335) becomes accessible to extracellular MTSET, while three residues (338, 340 and 344) in a previously intracellularly accessible segment of S4 become inaccessible to intracellular MTSET. In addition, our results show that one residue (338) is intracellularly accessible at hyperpolarized potentials and extracellularly accessible at depolarized potentials, suggesting that S4 moves outward in response to a depolarization. In the closed state (at depolarized potentials), at least six residues (339–344) are inaccessible (Fig. 4a). This movement of S4 strongly argues for S4 as the voltage sensor in spHCN channels. In addition, when we prevented S4 from moving outward by modifying cysteines in S4 with MTSET, spHCN channels were prevented from closing, supporting our hypothesis that S4 is the voltage sensor in spHCN channels and that S4 movement triggers the opening and closing of the activation gate in spHCN channels.

Our results show that the transmembrane movement of the voltage sensor S4 is conserved between the depolarization-activated Kv channels and the hyperpolarization-activated HCN channels. Figure 4a shows the transmembrane positions of S4 in an spHCN channel and a Shaker K^+ channel, based on the accessibility results from earlier studies of the Shaker K channel^{11,17} and from the present investigation of the spHCN channel. The similarities between these two channels in cysteine accessibilities of S4 residues suggest that the role of S4 as voltage sensor and the movement of S4 are conserved between the HCN channels and the Kv channels. In addition, the fact that MTSET modification of an S4 residue (338C) eliminates most of the gating currents in spHCN channels supports our conclusion that S4 is the main voltage sensor in spHCN channels.

spHCN channels have been shown to have a gate towards the cytosolic end of the pore that is opened by hyperpolarization and closed by depolarization^{20,22}. It has been suggested that this gate is similar to the activation gate in the Shaker K^+ channel^{23,24}, although

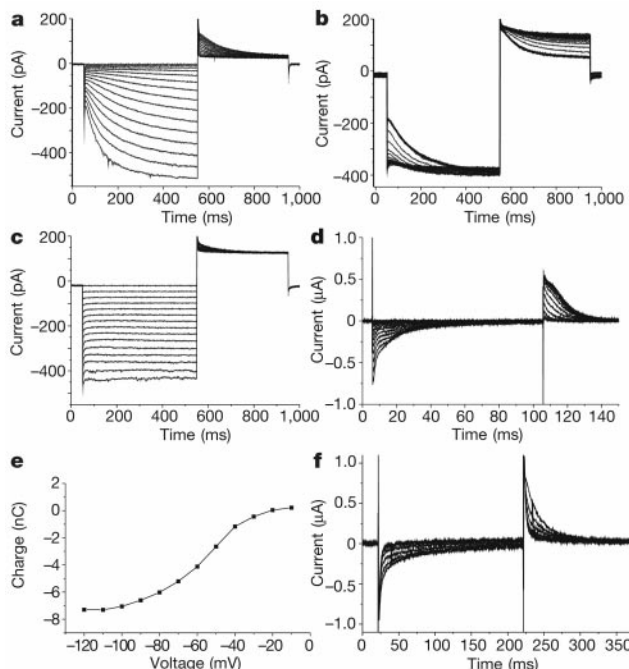


Figure 3 MTSET modification locks activation gate and S4. **a–c**, Currents through L340C (**a**) before, (**b**) during, and (**c**) after internal MTSET. Voltage steps from -10 mV to -150 mV in **a** and **c**. Voltage step to -120 mV in **b** for all traces. Holding potential, -10 mV. Tail potential, $+50$ mV. **d**, Gating currents in response to voltage steps ranging from 0 to -110 mV, from a holding potential of -10 mV. Tail potential, $+50$ mV. **e**, Voltage dependence of the gating-charge movement $Q(V)$, fitted with a single Boltzmann curve: $Q(V) = Q_{\text{max}} / (1 + \exp[-e_0 z(V - V_{1/2}) / kT])$. $V_{1/2} = -58 \pm 6$ mV, $z = 1.9 \pm 0.1$ ($n = 4$). **f**, Application of 5-mM extracellular MTSET reduced the gating currents of S338C channels (20 s of MTSET application between traces). Voltage step to -100 mV from 0 mV for all traces. Tail potential, $+50$ mV.

Table 1 Modification rates for cysteines in S4 of spHCN channels

	Extracellular		Intracellular	
	k at -100 mV ($\text{s}^{-1} \text{M}^{-1}$)	k at $+50$ mV ($\text{s}^{-1} \text{M}^{-1}$)	k at -100 mV ($\text{s}^{-1} \text{M}^{-1}$)	k at 0 mV ($\text{s}^{-1} \text{M}^{-1}$)
R326C	370 ± 130 (3)	260 ± 110 (3)	No effect (2)	
R332C	500 ± 220 (4)	$1,450 \pm 550$ (4)	No effect (2)	
K335C	<16 (5)	220 ± 140 (5)	No effect (4)	
S338C	2.5 ± 0.5 (4)	24.5 ± 3.0 (3)	$7,400 \pm 2,300$ (6)	53 ± 20 (8)
L340C	No effect (3)		$2,400 \pm 1,100$ (4)	100 ± 50 (4)
R344C	No effect (3)		$5,700 \pm 1,300$ (6)	76 ± 32 (3)

Values given as mean $\pm$ s.e.m. (n). k is the second-order rate constant for MTSET modification of the specified channels at the indicated potentials. 'No effect' means that 1 mM MTSET for 2 min affected the current by less than 10%; thus, $k < 1 \text{ s}^{-1} \text{M}^{-1}$.

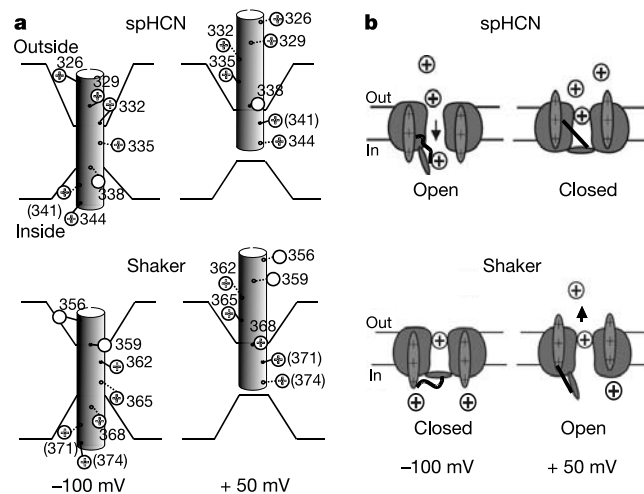


Figure 4 S4 movement is conserved between Shaker and spHCN channels. **a**, Sketch showing the implied transmembrane movement of S4 in Shaker and HCN channels, as suggested by the cysteine accessibility data in Shaker^{11,17} and spHCN channels, interpreted using the helical-screw model for S4 motion²⁹. **b**, Sketch showing that the outward movement of S4 opens the Shaker K⁺ channels, but closes the spHCN channels,

the activation gate in Shaker is opened by depolarization. Our results suggest that the activation gate in HCN channels is opened by the inward movement of S4 in response to hyperpolarizations and is closed by the outward movement of S4 in response to depolarizations. Although our results show that S4 is probably the main voltage sensor in HCN channels, it is still not known how S4 movement causes the activation gate to open. Both the S4 movement^{11,17} and the activation gate^{20,22–24} appear to be conserved between the hyperpolarization-activated HCN channels and the depolarization-activated Kv (Shaker) channels. Because HCN and Kv channels open in response to voltage steps with opposite polarity, we propose that the coupling mechanism between the voltage sensor and the activation gate in HCN channels is different from that in Kv channels (Fig. 4b). □

Methods

The spHCN channel

The experiments were performed on the spHCN channel—an ion channel from sea urchin (*Strongylocentrotus purpuratus*)⁹ that belongs to the family of HCN channels—expressed in *Xenopus* oocytes. spHCN activates at less negative voltages than the mammalian counterpart. This allowed us to test for S4 movement without having to clamp the oocytes to extreme negative potentials, which would have activated endogenous channels in the *Xenopus* oocytes and contaminated our current measurements. The spHCN channel displays a cAMP-dependent inactivation not present in the mammalian HCN channels⁹. This inactivation can be removed by increasing the cAMP levels on the cytosolic side of the spHCN channel, making the currents through the channel appear similar to the currents through the mammalian HCN channels^{9,20,22}.

Molecular biology

Site-directed mutagenesis, *in vitro* synthesis of RNA, and RNA injection into *Xenopus laevis* oocytes were performed as described previously²⁵.

Electrophysiology and solutions

The cysteine-substituted channels expressed in *Xenopus* oocytes were recorded with a two-electrode voltage-clamp technique for the extracellular accessibility studies, with the patch-clamp technique (macropatches) for the intracellular accessibility studies, and with the cut-open oocyte voltage-clamp technique for the gating current measurements. Microelectrodes were made from borosilicate glass, and filled with a 3 M KCl solution (for two-electrode and for cut-open oocyte experiments) or an extracellular solution (for patch-clamp experiments). The resulting resistance was 0.5–2.0 MΩ for two-electrode, 0.2–0.6 MΩ for cut-open oocyte, and 3–7 MΩ for patch-clamp experiments. Analogue capacitance compensation was used but not leak compensation. The currents were low-pass filtered at 1 or 3 kHz. All experiments were performed at room temperature (20–23°C).

implying that there is a different coupling mechanism between S4 and the activation gate in HCN channels than in Kv channels. For simplicity, the difference in coupling mechanism between Kv and HCN channels is shown as a difference in the attachment site of a linker between S4 and the activation gate. The true molecular nature of the coupling mechanism is unknown for both channels.

Two-electrode voltage-clamp technique

The whole-cell ion currents were measured with the two-electrode voltage-clamp technique (CA-1B amplifier, Dagan Corp.). In the bath, a 100-mM K (100-K) solution was used (in mM): 89 KCl, 15 HEPES, 0.4 CaCl₂ and 0.8 MgCl₂; KOH was added to adjust pH to 7.4, yielding a final K concentration of about 100 mM. The recordings were done without any manipulation of the intracellular cAMP concentrations. The endogenous cAMP levels in *Xenopus* oocytes has been found^{26,27} to be 1–4 μM, which is high enough to remove most of the cAMP-dependent inactivation of spHCN channels ($K_d = 0.74$ μM; ref. 9). Indeed, application of 10 μM forskolin and 1 mM 3-isobutyl-1-methylxanthine for 20 min increased the amplitude of the spHCN currents by <10% ($n = 4$).

Patch voltage-clamp technique

The macro patch currents were measured with the patch-clamp technique (Axopatch 200B, Axon Instruments). For the extracellular pipette solution, we used 100-K described above. 100 μM cAMP was included in the intracellular solution (in mM: 100 KCl, 0.5 MgCl₂, 0.1 CaCl₂, 1 EGTA and 10 HEPES) and was perfused onto the inside-out patches, to ensure that the spHCN currents were non-inactivating and were large enough to be easily measured.

Cut-open oocyte voltage-clamp technique

The gating currents were measured with the cut-open oocyte technique²⁸ on oocytes injected with high levels of cRNA (100 ng versus 1 ng for TEVC recordings). The solution in the (extracellular) top pool and the guard pool was composed of (in mM): 107 KOH, 107 methanesulphonic acid, 10 HEPES and 2 CaCl₂. The solution in the (intracellular) lower pool was (in mM): 110 KOH, 110 methanesulphonic acid, 10 HEPES and 0.1 EGTA. After RNA injection, the oocytes were incubated for >3 d with 1 mM ZD7288 (Tocris Cookson Inc.), the specific HCN channel blocker, to ensure that all HCN channels were blocked. Incubating the oocyte with a high concentration of ZD7288 increased the survival of the oocytes expressing high levels of channels. The extracellular recording solution also contained 1 mM ZD7288, and in some experiments, 1–10 mM Cs⁺ was also added, to ensure that all ionic currents were blocked. However, the Cs⁺ ions did not affect the currents, suggesting that 1 mM ZD7288 was enough to block the ion current completely. Cancellation of endogenous, linear capacitance transients was performed at +20 mV, to avoid activating the gating currents of spHCN, which move at potentials more negative than 0 mV (see Fig. 3e). The recordings were made without any leak compensation, P/4 compensation³⁰, or averaging. Control experiments were conducted with uninjected oocytes that underwent the same treatment as the RNA-injected oocytes. A comparison of the RNA-injected oocytes with the uninjected oocytes ensured that the transient currents we measured were from the expressed HCN channels and not due to transient currents endogenous to the oocytes. For the quantitative analysis of the experiments, we used off-line digital leakage compensation (that is, baseline subtraction of the current at the end of the pulse).

Methane thio sulphonate modification

The cysteines were tested for state-dependent, external or internal accessibility with the membrane-impermeable cysteine reagent MTSET in the open and closed states of the channel as described previously^{11,17}. Closed-channel modification was measured at +50 mV in TEVC recordings and 0 mV in excised patch recordings because the $G(V)$ curves of the spHCN channels measured in whole oocytes are shifted by 20–30 mV compared to $G(V)$ s measured in excised patches⁹. The cAMP modulation of the current (intracellular cAMP increases the amplitude of the current) was used to test the ability of

the perfusion system to quickly change the solution at the intracellular side of the excised patches¹¹. The MTSET modification of the gating currents was measured using the TEVC technique to ensure good perfusion of the oocyte.

Received 17 May; accepted 12 July 2002; doi:10.1038/nature01038.

1. Santoro, B. & Tibbs, G. R. The HCN gene family: molecular basis of the hyperpolarization-activated pacemaker channels. *Ann. NY Acad. Sci.* **868**, 741–764 (1999).
2. DiFrancesco, D. Pacemaker mechanisms in cardiac tissue. *Annu. Rev. Physiol.* **55**, 455–472 (1993).
3. Pape, H. C. Queer current and pacemaker: the hyperpolarization-activated cation current in neurons. *Annu. Rev. Physiol.* **58**, 299–327 (1996).
4. Magee, J. C. Dendritic hyperpolarization-activated currents modify the integrative properties of hippocampal CA1 pyramidal neurons. *J. Neurosci.* **18**, 7613–7624 (1998).
5. Magee, J. C. & Carruth, M. Dendritic voltage-gated ion channels regulate the action potential firing mode of hippocampal CA1 pyramidal neurons. *J. Neurophysiol.* **82**, 1895–1901 (1999).
6. Beaumont, V. & Zucker, R. S. Enhancement of synaptic transmission by cyclic AMP modulation of presynaptic Ih channels. *Nature Neurosci.* **3**, 133–141 (2000).
7. Mellor, J., Nicoll, R. A. & Schmitz, D. Mediation of hippocampal mossy fiber long-term potentiation by presynaptic Ih channels. *Science* **295**, 143–147 (2002).
8. Kaupp, U. B. & Seifert, R. Molecular diversity of pacemaker ion channels. *Annu. Rev. Physiol.* **63**, 235–257 (2001).
9. Gauss, R., Seifert, R. & Kaupp, U. B. Molecular identification of a hyperpolarization-activated channel in sea urchin sperm. *Nature* **393**, 583–587 (1998).
10. Yang, N. & Horn, R. Evidence for voltage-dependent S4 movement in sodium channels. *Neuron* **15**, 213–218 (1995).
11. Larsson, H. P., Baker, O. S., Dhillon, D. S. & Isacoff, E. Y. Transmembrane movement of the shaker K⁺ channel S4. *Neuron* **16**, 387–397 (1996).
12. Yang, N., George, A. L. Jr & Horn, R. Molecular basis of charge movement in voltage-gated sodium channels. *Neuron* **16**, 113–122 (1996).
13. Aggarwal, S. K. & MacKinnon, R. Contribution of the S4 segment to gating charge in the Shaker K⁺ channel. *Neuron* **16**, 1169–1177 (1996).
14. Seoh, S. A., Sigg, D., Papazian, D. M. & Bezanilla, F. Voltage-sensing residues in the S2 and S4 segments of the Shaker K⁺ channel. *Neuron* **16**, 1159–1167 (1996).
15. Yusaf, S. P., Wray, D. & Sivaprasadarao, A. Measurement of the movement of the S4 segment during the activation of a voltage-gated potassium channel. *Pflügers Arch.* **433**, 91–97 (1996).
16. Mannuzzu, L. M., Moronne, M. M. & Isacoff, E. Y. Direct physical measure of conformational rearrangement underlying potassium channel gating. *Science* **271**, 213–216 (1996).
17. Baker, O. S., Larsson, H. P., Mannuzzu, L. M. & Isacoff, E. Y. Three transmembrane conformations and sequence-dependent displacement of the S4 domain in shaker K⁺ channel gating. *Neuron* **20**, 1283–1294 (1998).
18. Vaca, L. *et al.* Mutations in the S4 domain of a pacemaker channel alter its voltage dependence. *FEBS Lett.* **479**, 35–40 (2000).
19. Chen, J., Mitcheson, J. S., Lin, M. & Sanguinetti, M. C. Functional roles of charged residues in the putative voltage sensor of the HCN2 pacemaker channel. *J. Biol. Chem.* **275**, 36465–36471 (2000).
20. Shin, K., Rothberg, B. & Yellen, G. Blocker state dependence and trapping in hyperpolarization-activated cation channels. Evidence for an intracellular activation gate. *J. Gen. Physiol.* **117**, 91–102 (2001).
21. Horn, R., Ding, S. & Gruber, H. J. Immobilizing the moving parts of voltage-gated ion channels. *J. Gen. Physiol.* **116**, 461–476 (2000).
22. Rothberg, B. S., Shin, K. S., Phale, P. S. & Yellen, G. Voltage-controlled gating at the intracellular entrance to a hyperpolarization-activated cation channel. *J. Gen. Physiol.* **119**, 83–91 (2002).
23. Liu, Y., Holmgren, M., Jurman, M. E. & Yellen, G. Gated access to the pore of a voltage-dependent K⁺ channel. *Neuron* **19**, 175–184 (1997).
24. del Camino, D., Holmgren, M., Liu, Y. & Yellen, G. Blocker protection in the pore of a voltage-gated K⁺ channel and its structural implications. *Nature* **403**, 321–325 (2000).
25. Larsson, H. P. & Elinder, F. A conserved glutamate is important for slow inactivation in K⁺ channels. *Neuron* **27**, 573–583 (2000).
26. Allende, C. C., Bravo, R. & Allende, J. E. Comparison of in vivo and in vitro properties of cyclic adenosine 3':5'-monophosphate phosphodiesterase of amphibian oocytes. *J. Biol. Chem.* **252**, 4662–4666 (1977).
27. Maller, J. L., Butcher, F. R. & Krebs, E. G. Early effect of progesterone on levels of cyclic adenosine 3':5'-monophosphate in *Xenopus* oocytes. *J. Biol. Chem.* **254**, 579–582 (1979).
28. Taglialatela, M., Toro, L. & Stefani, E. Novel voltage clamp to record small, fast currents from ion channels expressed in *Xenopus* oocytes. *Biophys. J.* **61**, 78–82 (1992).
29. Keynes, R. D. & Elinder, F. The screw-helical voltage gating of ion channels. *Proc. R. Soc. Lond. B* **266**, 843–852 (1999).
30. Bezanilla, F. & Armstrong, C. M. Inactivation of the sodium channel. I. Sodium current experiments. *J. Gen. Physiol.* **70**, 549–566 (1977).

Supplementary Information accompanies the paper on Nature's website
(<http://www.nature.com/nature>).

Acknowledgements We thank U. B. Kaupp for providing the spHCN clone; A. Broomand, H. Sjölander and S. Pandey for help with some of the recordings; P. Århem for comments on the manuscript; and S. Oster for editing the manuscript. This study was supported by the Swedish Research Council, Åke Wibergs Stiftelse, Magn. Bergvalls Stiftelse, The Swedish Society of Medicine, and Jeansson's Stiftelser and start-up funds from Oregon Health & Science University. F.E. has a junior research position at the Swedish Research Council. R.M. is supported by a pre-doctoral grant from the National Network in Neuroscience in Sweden.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to H.P.L.
(e-mail: larssonp@ohsu.edu).

Dec1 and Dec2 are regulators of the mammalian molecular clock

Sato Honma*, Takeshi Kawamoto†, Yumiko Takagi*, Katsumi Fujimoto†, Fuyuki Sato†, Mitsuhide Noshiro†, Yukio Kato† & Ken-ichi Honma*

* Department of Physiology, Hokkaido University Graduate School of Medicine, Sapporo 060-8638, Japan

† Department of Dental and Medical Biochemistry, Hiroshima University Graduate School of Biomedical Sciences, Hiroshima 734-8553, Japan

The circadian rhythms in mammals are regulated by a pacemaker located in the suprachiasmatic nucleus of the hypothalamus^{1,2}. Four clock-gene families have been found to be involved in a transcription–translation feedback loop that generates the circadian rhythm at the intracellular level³. The proteins Clock and Bmal1 form a heterodimer which activates the transcription of the *Per* gene from the E-box elements in its promoter region^{4,5}. Protein products of *Per* act together with *Cry* proteins to inhibit *Per* transcription^{6,7}, thus closing the autoregulatory feedback loop. We found that Dec1 and Dec2, basic helix–loop–helix transcription factors, repressed Clock/Bmal1-induced transactivation of the mouse *Per1* promoter through direct protein–protein interactions with Bmal1 and/or competition for E-box elements. Dec1 and Dec2 are expressed in the suprachiasmatic nucleus in a circadian fashion, with a peak in the subjective day. A brief light pulse induced Dec1 but not Dec2 expression in the suprachiasmatic nucleus in a phase-dependent manner. Dec1 and Dec2 are regulators of the mammalian molecular clock, and form a fifth clock-gene family.

Circadian rhythms are evolutionally conserved from bacteria to humans⁸. All clock genes so far identified in mammals are components of an autoregulatory transcription–translation feedback loop³. Gene products of *Clock* and *Bmal1* act as positive components, whereas those of the *Per* and *Cry* genes act as negative ones^{4–7}. This molecular feedback loop is widely accepted to be the intracellular machinery for circadian rhythm generation, but the model is insufficient to explain the stability and precision of circadian rhythms in living cells. Interlocked multiple feedback loops are considered as additional components to stabilize the molecular machinery^{3,9,10}.

DEC1 (*Stra13/Sharp2/BHLHB2*) and *DEC2* (*Sharp1/BHLHB3*) encode basic helix–loop–helix (bHLH) transcription factors related to rat *Hes*, *Drosophila* *Hairy*, and Enhancer of split (E(spl))^{11–13}. Four of us (T.K., K.F., M.N., and Y.K.) cloned complementary DNA of human *DEC1*¹¹, a cyclic AMP inducible gene product, in embryonic chondrocytes and cloned cDNAs of *DEC2* in human, mouse and rat, by expressed sequence tag (EST) searching for similar members to *DEC1*¹⁴. In contrast to other members of *Hes/Hairy/E(spl)* subgroups, *DEC1* and *DEC2* lack a carboxy-terminal 'WRPW domain' to which co-repressors bind^{11,14}. Circadian expression in peripheral tissues of *Dec1* and *Dec2* (M.N., manuscript in preparation), and reports that related proteins can bind to E-box elements^{15,16}, prompted us to study the roles of Decs in the mammalian circadian system.

Both Dec1 and Dec2 suppressed Clock/Bmal1-induced activation of *Per1* promoter in a dose-dependent manner as reported by a luciferase assay (Fig. 1a). Firefly reporter plasmid (E54-TK) was constructed by inserting a 5' flanking region of mouse *Per1* containing 3 E-box elements, CACGTG, upstream of the TK promoter and these were transfected to NIH3T3 cells. The luciferase activity was markedly increased (125-fold) in the presence of *Clock* and *Bmal1* plasmid (50 ng per well). Co-transfection of the same amount of *Dec1* or *Dec2* plasmid suppressed the promoter activity down to $13.3 \pm 1.0\%$ (mean $\pm$ s.e.) and to $5.9 \pm 0.5\%$, respectively. Even

with one-eighth of the amount of plasmid, Dec1 suppressed the promoter activity to $72.5 \pm 4.5\%$, and Dec2, to $14.6 \pm 1.1\%$. The suppressing effects of Dec2 overwhelmed those of Per1, Per2, CRY1 and CRY2 at the same dosage. In our luciferase assay, CRY1 and CRY2 inhibited the promoter activity more strongly than did Per1 or Per2, confirming a previous result⁶. The suppression by Dec1 was comparable to that by CRY1 and CRY2.

Dec1 and Dec2 are not general inhibitors of transcription. Hypoxia inducible factor-1 α (HIF-1 α) is a bHLH-PAS transcription factor like Clock and Bmal1, and activates transcription of targets such as the erythropoietin gene through hypoxia-response elements (HRE)¹⁷. We examined the effects of Dec1 or Dec2 cotransfection on a HIF-1 α -induced increase of the HRE reporter activity, and found that they did not significantly change the reporter activity ($75.3 \pm 5.3\%$ and $114.9 \pm 9.6\%$, respectively)—which was increased by HIF-1 α (20.4-fold).

Gel retardation assays show that both mouse Dec1 and mouse Dec2 bound to mouse *Per1* proximal E-box elements (Fig. 1b). The binding was specific to an E-box, because excess dose of unlabelled E-box but not mutant E-box inhibited the binding. Thus, Decs may suppress Clock/Bmal1 induced transactivation by competing for E-box binding upstream of mouse *Per1*. Nevertheless, a yeast

two-hybrid assay revealed that both Dec1 and Dec2 bound stereospecifically to Bmal1 (Fig. 1c). Dec1 and Dec2 lacking the C-terminal regions were able to bind to Bmal1 as did the full-length Decs, but those lacking the amino terminal failed to bind, suggesting the bHLH is the responsible site. Dec1 and Dec2 with the N-terminal regions did not bind to Clock. In this respect, Decs are different from E4BP4, a basic leucine zipper transcription factor which represses *Per1* transcription by competitively binding to the albumin D-site binding protein (DBP)-binding consensus at the 5' upstream of *Per1*¹⁸. Dec1 and Dec2 exert their repressing effects on Clock/Bmal1-induced transactivation through competition for E-boxes with Clock/Bmal1 and/or direct protein–protein interactions with Bmal1.

Dec1 and *Dec2* are expressed in the suprachiasmatic nucleus of the hypothalamus (SCN) in a circadian fashion (Fig. 2). Robust circadian rhythms ($P < 0.001$) were detected in both gene expression under continuous darkness (DD) as well as under light–dark conditions (LD) with peaks either early (*Dec1*) or in the middle (*Dec2*) of subjective day (Fig. 2a, b). Different light conditions did not change the circadian profiles. Compared with other brain areas, the hybridization signals in the SCN were weak for *Dec1*, but strong for *Dec2* (Fig. 2c–f).

The circadian rhythms are phase-shifted by a light pulse, which is an important mechanism for the entrainment of the circadian clock

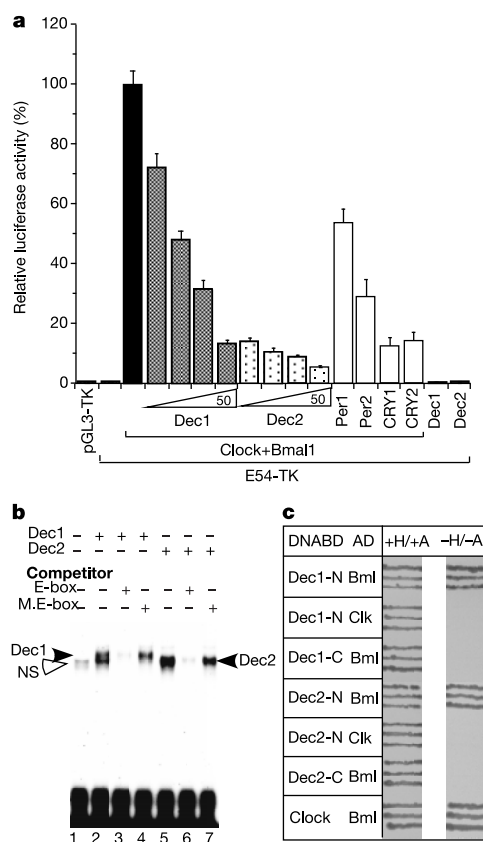


Figure 1 Repression of the Clock/Bmal1-induced E54-TK reporter activity by Dec1 and Dec2, and mechanisms of the repression. **a**, Dose-dependent repression of the Clock/Bmal1-induced E54-TK reporter activity by Dec1 and Dec2, compared with repression by Per1, Per2, CRY1 and CRY2. A relative luciferase activity was calculated as per cent values of the mean luminescence levels of E54-TK with Clock/Bmal1, and expressed in mean $\pm$ s.e. Results represent two independent experiments with $n = 4$ of each condition. **b**, Specific binding of Dec1 and Dec2 to the *Per1* proximal E-box by gel retardation assay. Arrow heads, binding bands with Dec1 and Dec2; open arrow head, non-specific band. **c**, Yeast two-hybrid assay showing that N-terminal (-N) but not C-terminal (-C) regions of Dec1 and Dec2 bind to Bmal1 (Bml). Clk, Clock; +H/+A, -Leu/-Trp SD plate; -H/-A, -Leu/-Trp/-His/-Ade SD plate.

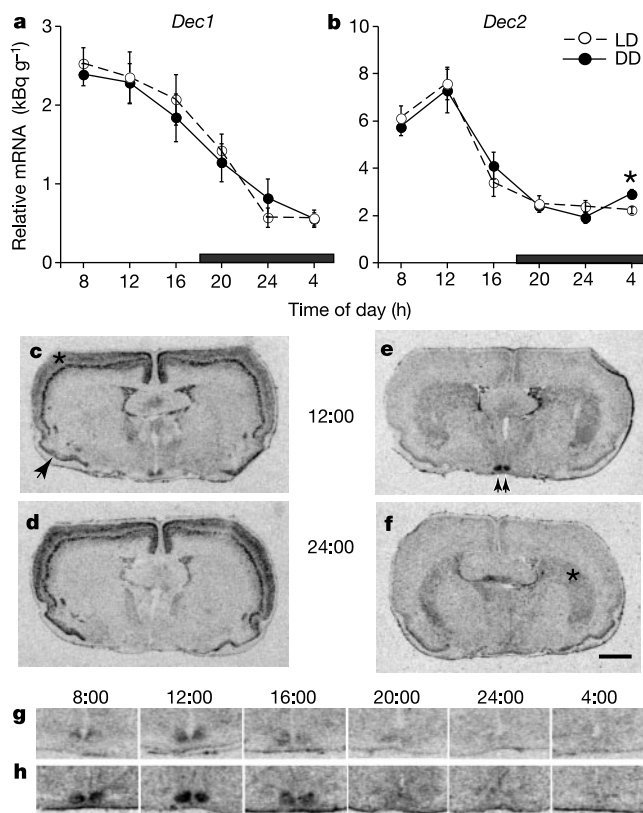


Figure 2 Circadian profiles of *Dec1* and *Dec2* expression in the SCN. *Dec1* (**a**) and *Dec2* (**b**) expression levels under light–dark (LD; open circles) and continuous darkness (DD; filled circles) (mean $\pm$ s.e.). Dark horizontal bars on the abscissa, dark period under LD cycle. Asterisk, $P < 0.05$ versus LD (post hoc test). Representative autoradiographs show *Dec1* (**c**, **d**) and *Dec2* (**e**, **f**) signals of the coronal brain sections including the SCN on the third day of DD at 12:00 (**c**, **e**) and 24:00 (**d**, **f**). Scale bar (**f**), 2 mm. Sampled areas of the parietal (star) and piriform cortices (an arrow) in **c**, SCN (arrows in **e**), and caudate putamen (star in **f**) are indicated. Higher magnification autoradiographs show *Dec1* (**g**) and *Dec2* (**h**) signals in the SCN at six times of day. Scale bar, 1 mm.

to 24-h light–dark environments. Light signals reach the SCN via the retinohypothalamic tract¹⁹, and shift the circadian clock—the amount and direction (advance or delay) of the shift depends on the phase of the light pulse²⁰. We examined light responsiveness of *Dec1* and *Dec2* gene expression in rats kept under DD for 3 d and subsequently exposed to a single light pulse of 30 min. *Dec1* and *Dec2* responded differently to the light pulse (Fig. 3). The light pulses given at 20:00, 24:00 and 4:00 significantly increased *Dec1* expression, whereas the same light pulses failed to induce *Dec2* expression. In contrast, the light pulse at 24:00 slightly but significantly decreased *Dec2* expression. Light pulses at other times did not change the level of gene expression for either *Dec1* or *Dec2*. Thus, the light responsiveness of *Dec1* expression was phase-dependent, and corresponded to the light-sensitive portion of a phase response curve²⁰ (Fig. 3a, c).

The circadian profiles of *Dec1* and *Dec2* expression were also examined in brain structures other than the SCN. Gene expression for *Dec1* and *Dec2* was different in different brain areas (Fig. 2c–f and Fig. 4a, b). *Dec1* was expressed strongly in the cerebral cortex, especially in the fifth layer, thalamus, and superior colliculus. On the other hand, *Dec2* expression was robust in the caudate putamen, pineal gland, and granular cell layer of the cerebellum. Both *Dec1* and *Dec2* were expressed in the olfactory bulb, piriform cortex, hippocampus and hypothalamic nuclei. Significant circadian rhythms were detected in the parietal and piriform cortices for *Dec1*, and in the piriform cortex and caudate putamen for *Dec2* under both LD and DD (Fig. 4c–f). None of them, however, were as robust as in the SCN. The circadian rhythms in these brain areas were almost 180° out of phase to those in the SCN. Phase reversal of the circadian rhythm outside the SCN has been observed in clock genes such as *Per1*, *Per2* and *Bmal1*^{21,22} and may be related to behavioural rhythms²². Rhythmic expression of *Dec1* and *Dec2* outside the SCN may reflect their roles in the interacting molecular feedback loops of peripheral clocks²³.

Recently, two interacting molecular loops were suggested to be involved in the clock mechanism in mammals⁹. Together with a core feedback loop of *Per* transcription with positive (Clock, *Bmal1*) and negative components (*Per1*, *Per2*, *Cry1*, *Cry2*), an additional feedback loop of *Bmal1* was suggested to be involved in such a way that

Bmal1 expression was enhanced by *Cry1*, *Cry2* and *Per2*^{9,10}. In the present study, Decs are suggested as strong negative components of the core loop. As *Dec* expression is up-regulated by *Clock/Bmal1* heterodimers (T.K., manuscript in preparation), Decs seem to be important components of an additional autoregulatory feedback loop which interlocks with the core feedback loop of circadian oscillation. Specifically, *Per* proteins activate the production of *Bmal1*^{9,10}, which in turn stimulates *Per* and *Dec* transcription. Decs suppress the transactivation of *Per* gene by *Clock/Bmal1*, thus closing the loop. A physiological role for *Dec* proteins in the molecular clock mechanism is supported by the timing of their expressions in the SCN. High expression in the middle of the subjective day matches the time of maximum *Clock/Bmal1* transcriptional activity²⁴. A recent study showed that most of the cyclic transcripts were tissue-specific and only a limited number of genes were rhythmically expressed both in the SCN and liver²⁴, suggesting a key role for these genes in the molecular clock mechanisms. Therefore, rhythmic expression of *Decs* in many peripheral tissues including the liver implies physiological significance (M.N., manuscript in preparation).

A single light pulse enhanced *Dec1* but not *Dec2* expression in a phase-dependent manner, which was similar to *Per1*^{25,26}. The light-induced *Dec1* expression is not due to a light-induced increase in *Per* or *Bmal1* proteins but is probably due to a direct action of light. Entraining light signals in mammals are mediated by the retinohypothalamic tract¹⁹, a major neurotransmitter of which is glutamate.

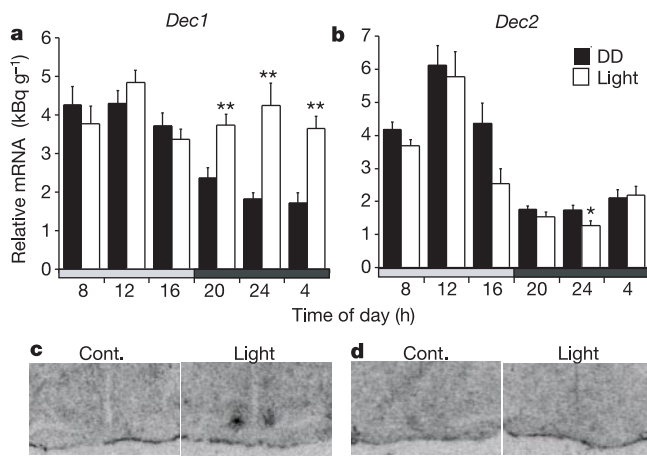


Figure 3 Response of *Dec1* and *Dec2* expression to 30 min light. Circadian variations of the light responsiveness of *Dec1* (a) and *Dec2* (b) expression in the SCN examined on the third day of DD. White columns, relative mRNA levels at 1 h after the onset of 30-min light pulse. Black columns, levels of control rats without light exposure. Grey and black bars on the abscissa, subjective day and night, respectively. Data are mean $\pm$ s.e. of 5 animals. * $P < 0.05$, and ** $P < 0.01$ versus control. Representative autoradiographs show *Dec1* (c) and *Dec2* (d) signals before (Cont.) and 1 h after the light pulse (Light) at 24:00. Scale bar, 1 mm.

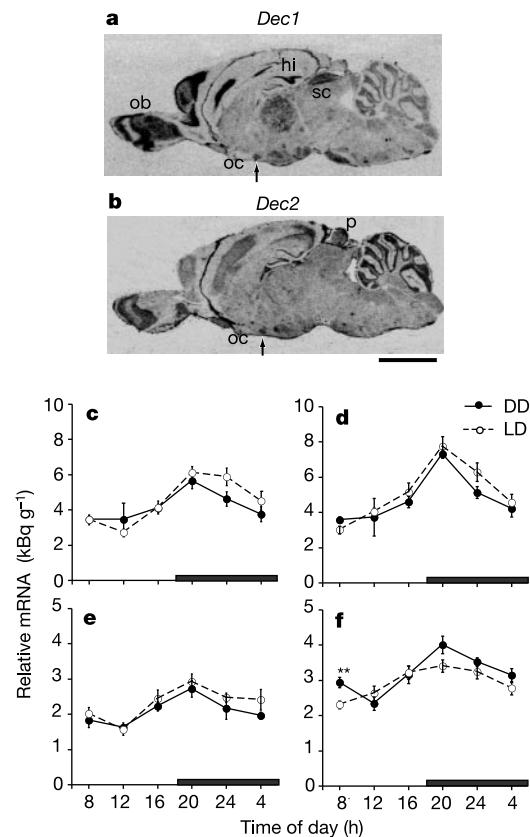


Figure 4 Circadian variation of *Dec1* and *Dec2* expression in the brain structures outside the SCN. Representative autoradiographs of *Dec1* (a) and *Dec2* (b) *in situ* hybridization in sagittal sections at 12:00 under LD cycle. Scale bar, 5 mm. Arrows indicate the SCN. hi, hippocampus; ob, olfactory bulb; oc, optic chiasm; p, pineal gland; and sc, superior colliculus. Circadian patterns of gene expression in the parietal (c) and piriform cortices (d) for *Dec1*, and the caudate-putamen (e) and piriform cortex (f) for *Dec2* under LD (open circles) and DD (filled circles). Dark horizontal bars on the abscissa, dark period under LD cycle. Data are mean $\pm$ s.e. of 5 animals. ** $P < 0.01$ versus LD (*post hoc* test).

Glutamate increases intracellular calcium, which in turn activates cAMP responsive element binding protein (CREB) in the SCN neurons²⁷. As happens for *Per1*²⁸, a binding site for CREB—calcium/cAMP responsive element (CRE)—is found in the upstream promoter region of *Dec1*¹⁵. *Per1* and *Dec1* seem to be the doors to the molecular-clock machinery. □

Methods

Transfection and luciferase assay

A firefly reporter plasmid, E54-TK, was constructed by connecting an E54 fragment into the *NheI*-*XhoI* site upstream of the TK promoter of the pGL3-TK vector²⁹ (a gift from M. Negishi and T. Sueyoshi). The E54 fragment consisted of three E-boxes within 2.0 kb of the 5' flanking region of the mouse *Per1* gene with their immediate flanking sequences linked together (TTTAGCCACGTGACAGTGAAGCACACGTGGGCCCTCAAGTCCACGTGCAGGGA) as reported previously⁴. NIH3T3 cells were plated (3.4×10^4 cells per well) in a 24-well plate 24 h before transfection. Reporter plasmid (E54-TK, 2 ng per well) and mouse *Clock* and mouse *Bmal1b* (each 50 ng per well) expression vectors were transfected with or without one of following vectors; human *DEC1*, mouse *Dec2*, mouse *Per1*, rat *Per2*, human *CRY1* or human *CRY2* (each 50 ng per well) using Lipofectamine PLUS (Gibco-BRL) according to the manufacturer's instruction. In a dose response experiment, *DEC1* and *Dec2* plasmids were added in doses of 50, 25, 12.5 and 6.125 ng per well. The total concentration of DNA was adjusted to 152.2 ng per well with pcDNA3 (Invitrogen). For the intrinsic standard, 0.2 ng of pRL-SV40 vector (Promega) was transfected to each well. Twenty-four hours after the transfection, cells were lysed with Cell Lysis Buffer (Promega), and renilla and firefly luciferase activities were examined with Dual Reporter Assay Reagent (Promega). Similarly, NIH3T3 cells were transfected with 4 ng of firefly reporter plasmid containing HRE, 50 ng of HIF-1 α (provided by L. Poellinger¹⁷) and 0.2 ng pRL-SV40 vector with or without 50 ng of *Dec1* or *Dec2* plasmid, and relative luciferase activity was measured ($n = 8$).

Expression constructs were made by subcloning full-length coding regions of the following genes into pcDNA3.1 (Invitrogen); mouse *Clock* (AF000998) and mouse *Bmal1b* (AB012601) were obtained by RT-PCR, mouse *Per1* was generously supplied by H. Tei²⁸, human *CRY1* (D84657), by M. Ikeda, and rat *Per2* (AB016532) and human *CRY2* (AB014558), by T. Nagase. As the AB014558 gene lacks 13 bp at the 5' end of the human *CRY2* coding region, the fragment was added by PCR according to the sequence of an EST clone (AL 040215).

Gel retardation assay

Mouse *Dec1* and *Dec2* were translated using the TNT quick coupled transcription/translation system (Promega). The double-stranded oligonucleotides of the *Per1* E-box (5'-cgcgCAAGTCCACGTGCAGGATcga-3') were end-labelled using [³²P]dCTP and DNA polymerase Klenow fragment. Synthesized mouse *Dec1* or *Dec2* was incubated with approximately 30,000 c.p.m. of ³²P-labelled probe for 10 min in 10 μ l of 10 mM Tris/HCl (pH 8.0), 0.5 mM dithiothreitol, 10% glycerol, 1 μ g of poly(dI-dC), 50 mM NaCl and 5 mM MgCl₂. In competition experiments, a 200-fold molar excess of unlabelled E-box or mutant E-box (5'-cgcgCAAGTCTcgtctCAGGATcga-3') was added to the reaction mixture. The mixtures were applied to 5% polyacrylamide gel electrophoresis in 12.5 mM Tris-HCl/125 mM glycine/1 mM EDTA buffer.

Yeast two-hybrid assay

cDNAs encoding full-length mouse *Bmal1*, mouse *Clock*, human *DEC1* and mouse *Dec2*, and those encoding N-terminal (4–139) and C-terminal (150–412) regions of human *DEC1* and N-terminal (1–196) and C-terminal (113–410) regions of mouse *Dec2* were cloned with pLP-GADT7AD, pLP-GBKT7DNA-BD, pGADT7AD or pGBKT7DNA-BD vector (Clontech). *Saccharomyces cerevisiae* AH109 was co-transfected with plasmids encoding either GAL4 DNA-BD or GAL4 AD fused protein. The transformants expressing two fused proteins were selected on plates of –Leu/–Trp SD medium. Three independent clones were assayed for transactivation of the HIS3 and ADE2 reporter genes on –Leu/–Trp/–His/–Ade SD plates.

Animal handling and tissue sampling

Male rats of Wistar strain were used at 2–2.5 months old. They were born and reared in our animal quarters, where environmental conditions were controlled (lights on from 06:00–18:00) as described elsewhere^{20–22}. In the first experiment, rats were decapitated under LD ($n = 30$) and on the third day in DD ($n = 30$) at 6 times of the day starting from 8:00. In the second experiment, rats ($n = 30$) kept under DD for 3 d were exposed to a light pulse of 300 lux for 30 min from one of six time points as in the first experiment. They were returned to DD and decapitated 1 h after the onset of the light pulse. Control rats ($n = 30$) were decapitated under DD without exposing to light. The brains were quickly removed and frozen as described previously^{21,22}.

Quantitative *in situ* hybridization

This was performed as described previously^{21,22}. Twelve serial sections containing the SCN in each rat were alternatively hybridized with *Dec1* and *Dec2* probes (5'-ACATACAAGCACAGCAGCGCACTACACACTCACTCTCACACA-3' and 5'-AAAGGGTAGAGGGTGGGTGGCAGAAAGGGTGAGAAAGGGGCAAAA-3', respectively) radiolabelled with ³⁵S-dATP (NEN). In addition, sagittal sections at the level of the SCN were saved in 3 brains obtained at 12:00 under LD. The hybridization signals were analysed using a

microcomputer imaging device (Imaging Research). Optical density of the structure of interest was calibrated by converting it to relative signal intensity (kBq g⁻¹) using the ¹⁴C-acrylic standards in the same cassette, and normalized by subtracting that of the corpus callosum in the same section as described previously²¹.

Statistics

Circadian rhythms were statistically analysed by one-way analysis of variance, ANOVA. Differences between two circadian rhythms were analysed by two-way ANOVA with the Bonferroni–Dunn *post hoc* test.

Received 11 July; accepted 29 August 2002; doi:10.1038/nature01123.

- Moore, R. Y. & Eichler, V. B. Loss of a circadian adrenal corticosterone rhythm following suprachiasmatic lesions in the rat. *Brain Res.* **42**, 201–206 (1972).
- Stephan, F. & Zucker, I. Circadian rhythms in drinking behavior and locomotor activity are eliminated by suprachiasmatic lesions. *Proc. Natl Acad. Sci. USA* **54**, 1521–1527 (1972).
- Reppert, S. M. & Weaver, D. R. Molecular analysis of mammalian circadian rhythms. *Annu. Rev. Physiol.* **63**, 647–676 (2001).
- Gekakis, N. *et al.* Role of the CLOCK protein in the mammalian circadian mechanism. *Science* **280**, 1564–1569 (1998).
- Darlington, T. K. *et al.* Closing the circadian loop: CLOCK-induced transcription of its own inhibitors per and tim. *Science* **280**, 1599–1600 (1998).
- Kume, K. *et al.* mCRY1 and mCRY2 are essential components of the negative limb of the circadian clock feedback loop. *Cell* **98**, 193–205 (1999).
- Yagita, K. *et al.* Dimerization and nuclear entry of mPER proteins in mammalian cells. *Gene Dev.* **14**, 1353–1363 (2000).
- Dunlap, J. C. Molecular bases for circadian clocks. *Cell* **96**, 271–290 (1999).
- Shearman, L. P. *et al.* Interacting molecular loops in the mammalian circadian clock. *Science* **288**, 1013–1019 (2000).
- Yu, W., Nomura, M. & Ikeda, M. Interacting feedback loop within the mammalian clock: BMAL1 is negatively autoregulated and upregulated by CRY1, CRY2, and PER2. *Biochem. Biophys. Res. Commun.* **290**, 933–941 (2002).
- Shen, M. *et al.* Molecular characterization of the novel basic helix-loop-helix protein DEC1 expressed in differentiated human embryo chondrocytes. *Biochem. Biophys. Res. Commun.* **236**, 294–298 (1997).
- Boudjelal, M. *et al.* Overexpression of Stra 13, a novel retinoic acid-inducible gene of the basic helix-loop-helix family, inhibits mesodermal and promotes neuronal differentiation of P19 cells. *Gene Dev.* **11**, 2052–2065 (1997).
- Rosner, M. J. *et al.* SHARPs: mammalian enhancer-of-split- and hairy-related proteins coupled to neuronal stimulation. *Mol. Cell. Neurosci.* **9**, 460–475 (1997).
- Fujimoto, K. *et al.* Molecular cloning and characterization of DEC2, a new member of basic helix-loop-helix proteins. *Biochem. Biophys. Res. Commun.* **280**, 164–171 (2001).
- Teramoto, M. *et al.* Gene structure and chromosomal location of a human bHLH transcriptional factor DEC1–Stra13–SHARP-2/BHLHB2. *J. Biochem. (Tokyo)* **129**, 391–396 (2001).
- Dear, T. N. *et al.* Identification of interaction partners for the basic-helix-loop-helix protein E47. *Oncogene* **14**, 891–898 (1997).
- Kallio, P. J. *et al.* Regulation of the hypoxia-inducible transcription factor 1 α by the ubiquitin-proteasome pathway. *J. Biol. Chem.* **274**, 6519–6525 (1999).
- Mitsui, S. *et al.* Antagonistic role of E4BP4 and PAR proteins in the circadian oscillatory mechanism. *Genes Dev.* **15**, 995–1006 (2001).
- Moore, R. Y. & Card, J. P. Visual pathways and the entrainment of circadian rhythms. *Ann. NY Acad. Sci.* **453**, 123–133 (1985).
- Honma, K., Honma, S. & Hiroshige, T. Response curve, free-running period, and activity time in circadian locomotor rhythm of rats. *Jpn J. Physiol.* **35**, 643–658 (1985).
- Honma, S. *et al.* Circadian oscillation of *BMAL1*, a partner of a mammalian clock gene *Clock* in rat suprachiasmatic nucleus. *Biochem. Biophys. Res. Commun.* **250**, 83–87 (1998).
- Masubuchi, S. *et al.* Clock genes outside the suprachiasmatic nucleus involved in manifestation of locomotor activity rhythm in rats. *Eur. J. Neurosci.* **12**, 4206–4214 (2000).
- Yamazaki, S. *et al.* Resetting central and peripheral circadian oscillators in transgenic rats. *Science* **288**, 682–685 (2000).
- Panda, S. *et al.* Coordinated transcription of key pathways in the mouse by the circadian clock. *Cell* **109**, 307–320 (2002).
- Shigeyoshi, Y. *et al.* Light-induced resetting of a mammalian circadian clock is associated with rapid induction of the *mPer1* transcript. *Cell* **91**, 1043–1053 (1997).
- Zylka, M. J., Shearman, L. P., Weaver, D. R. & Reppert, S. M. Three period homologs in mammals: different light responses in the suprachiasmatic circadian clock and oscillating transcripts outside of brain. *Neuron* **20**, 1103–1110 (1998).
- Travnickova-Bendova, Z., Cermakian, N., Reppert, S. M. & Sassone-Corsi, P. Bimodal regulation of mPeriod promoters by CREB-dependent signaling and CLOCK/BMAL1 activity. *Proc. Natl Acad. Sci. USA* **99**, 7728–7733 (2002).
- Hida, A. *et al.* The human and mouse *Period 1* genes: five well-conserved E-boxes additively contribute the enhancement of *mPer1* transcription. *Genomics* **65**, 224–233 (2000).
- Sueyoshi, T. *et al.* The repressed nuclear receptor CAR responds to phenobarbital in activating the human CYP2B6 gene. *J. Biol. Chem.* **274**, 6043–6046 (1999).

Acknowledgements This work was supported in part by Grants-in-Aid for Scientific Research and Special Coordination Funds for Promoting Science and Technology from the Ministry of Education, Culture, Sports, Science and Technology, the Japanese Government.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to S.H. (e-mail: sathonma@med.hokudai.ac.jp).

Direct observation of ligand recognition by T cells

Darrell J. Irvine^{*}, Marco A. Purbhoo, Michelle Krogsgaard & Mark M. Davis

The Department of Microbiology & Immunology and The Howard Hughes Medical Institute, Stanford University School of Medicine, Stanford, California 94305, USA

The activation of T cells through interaction of their T-cell receptors with antigenic peptide bound to major histocompatibility complex (MHC) on the surface of antigen presenting cells (APCs) is a crucial step in adaptive immunity. Here we use three-dimensional fluorescence microscopy to visualize individual peptide–I-E^k class II MHC complexes labelled with the phyco-biliprotein phycoerythrin in an effort to characterize T-cell sensitivity and the requirements for forming an immunological synapse^{1–3} in single cells. We show that T cells expressing the CD4 antigen respond with transient calcium signalling to even a single agonist peptide–MHC ligand, and that the organization of molecules in the contact zone of the T cell and APC takes on the characteristics of an immunological synapse when only about ten agonists are present. This sensitivity is highly dependant on CD4, because blocking this molecule with antibodies renders T cells unable to detect less than about 30 ligands.

Previous studies have estimated that T cells carrying the CD4 antigen can be stimulated by as few as 50–400 peptides on an average cell^{4–7}, whereas T cells carrying the CD8 antigen are thought to require from 10 to 50 ligands^{8,9}, and perhaps even fewer¹⁰, to initiate cytotoxicity. But these findings are fraught with uncertainties because, first, it is not clear what percentage of an APC's surface is surveyed by a T cell and, second, it is very likely that some APCs will have more peptides loaded on them than others. To observe specific peptide–MHC complexes directly at the interface between T cells and APCs, we pulsed APCs with biotinylated peptides and the resulting biotin–MCC–I-E^k complexes were labelled with saturating amounts of phycoerythrin-conjugated streptavidin (SAV-PE). In preliminary experiments, single SAV-PE molecules that were adsorbed to coverslips coated with poly-L-lysine and imaged in phosphate-buffered saline (PBS) were detectable by fluorescence microscopy as diffraction-limited spots with a mean signal-to-noise ratio of 11.4:1. In agreement with previous studies of single molecules^{11,12}, we found that peak intensities of single PE spots showed narrow distributions. When adsorbed to coverslips at higher densities, PE fluorescence spots with two or three times the intensity of putative single PE spots occurred with a frequency in agreement with Poisson statistics for the occurrence of overlapping pairs or triplets of stochastically placed molecules.

To be sure that we were able to detect all or most peptides on the cell surface and that these were bound to I-E^k molecules, we compared the number of complexes per cell detected by labelling with SAV-PE to that detected by a PE-conjugated monoclonal antibody (D4)⁷ specific for MCC–I-E^k (Fig. 1). The number of MCC–I-E^k complexes detected by SAV-PE versus the antibody against MCC–I-E^k was determined by Scatchard analysis of binding data obtained by fluorescence-activated cell sorting (FACS; Supplementary Information). FACS measurements were confirmed by examining labelled cells by three-dimensional digital microscopy (see below) and converting the total cell fluorescence detected to peptide number by measuring the fluorescence from single discrete PE molecules in the same sample. SAV-PE detected essentially

identical amounts of peptide as the peptide–MHC-specific antibody over a broad range of total peptide densities, and measurements by microscopy were also in close agreement with these estimates (Fig. 1). Because the antibody is bivalent for the biotinylated complexes and streptavidin is tetravalent for biotinylated complexes, this result could only be obtained if the use of a saturating amount of the detection reagent (that is, antibody or streptavidin) led to monovalent binding.

The observation of monovalent binding of the tetravalent PE tag to peptide–MHC by labelling with excess SAV-PE is also consistent with the kinetics of labelling. Using saturating concentrations of SAV-PE, labelling was complete in about 2 min at the low densities of total peptide used (data not shown). By contrast, the average time required for a SAV-PE-bound peptide–MHC to encounter a second complex was about 25 times greater (Supplementary Information). Thus, the peptides should be tagged with streptavidin well before dimerization or higher-order oligomerization of complexes can occur, even at higher peptide concentrations.

Labelling biotin–MCC–I-E^k complexes with SAV-PE caused no observable perturbation to the responsiveness of T cells to APCs in both short- and long-term assays. Calcium signalling and synapse formation were identical in labelled and unlabelled T-cell–APC conjugates, as assessed by microscopy, and labelled MCC–I-E^k was stable during microscopy. Proliferation and interleukin 2 (IL-2) secretion were also indistinguishable from the responses of T cells to APCs loaded with wild-type MCC peptide (Supplementary Information).

To assess T-cell sensitivity, we loaded T-cell blasts taken from 5C.C7 T-cell receptor (TCR) transgenic mice with the intracellular calcium indicator dye fura-2 (ref. 13), mixed them with labelled APCs, and imaged them by time-lapse three-dimensional digital microscopy. Figure 2 shows representative responses to defined numbers of MCC–I-E^k complexes at the interface. T cells were found to stop, form a tight interface with the APC, and increase intracellular calcium in response to even one MCC–I-E^k complex. These increases in calcium induced by single peptides generally decayed rapidly (Fig. 2a, b). By contrast, T cells responding to two agonist peptides showed a more stable calcium signal over the same time period (Fig. 2c and Supplementary Information), indicating that the single-ligand effects that we observed were not due to rare peptide dimerization events. Contact with ten or more peptides typically stimulated a sustained increase in calcium, with fura ratios remaining roughly twofold higher than initial levels for at least

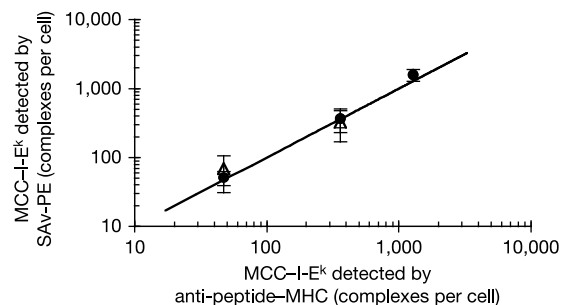


Figure 1 Detection of specific peptide–MHC complexes formed on the surface of APCs using streptavidin-PE versus a complex-specific antibody. Plotted are the numbers of MCC–I-E^k complexes on APCs detected by the PE-conjugated D4 antibody (D4-PE) specific for MCC–I-E^k in FACS versus the complex number detected by SAV-PE in FACS (filled circles) or SAV-PE in microscopy (open triangles). The line plots $x = y$, which is the relationship expected if SAV-PE detects the same number of complexes as the complex-specific antibody D4. Results are shown for a representative experiment with CH27 APCs; equivalent complex detection by SAV-PE and D4-PE was also found when the A20 cell line or primary splenocytes were used as APCs (not shown). Error bars indicate standard deviations.

^{*} Present address: Department of Materials Science & Engineering/Biological Engineering Division, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA.

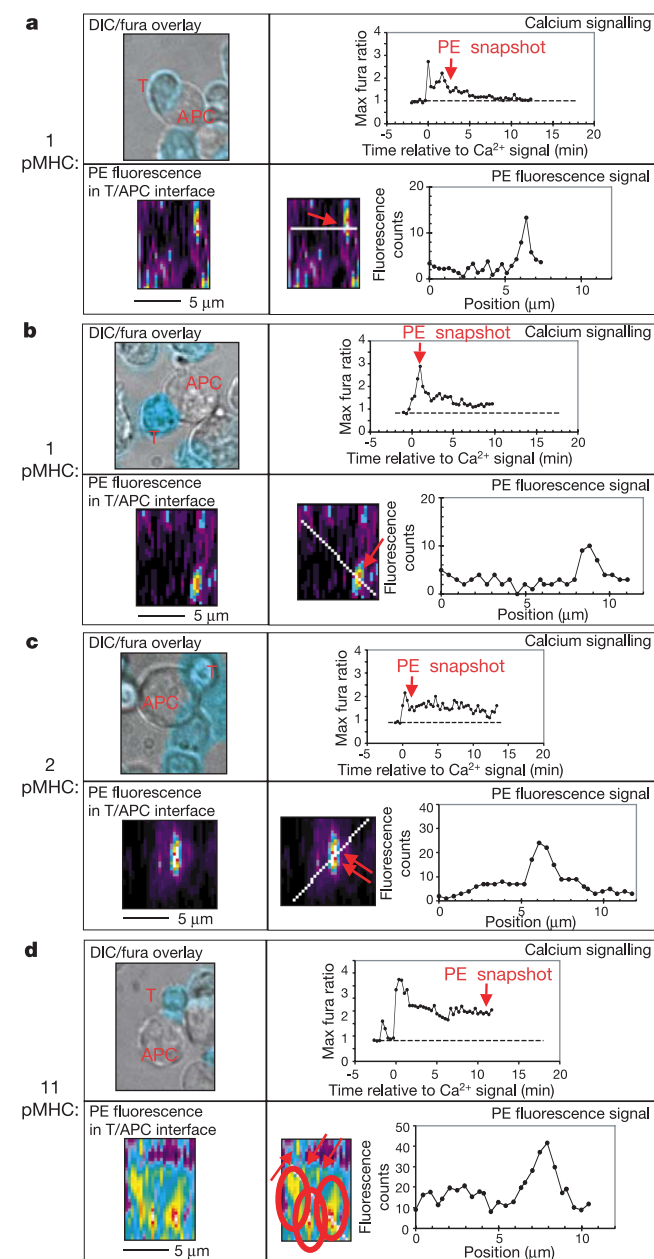


Figure 2 Calcium increases in T cells in response to defined numbers of peptide-MHC ligands. Shown are the distributions of individual MCC-I-E^k agonist ligands in the interface of responding cells (as revealed by fluorescent labelling) and the resultant calcium signals for T cells responding to 1 (a, b), 2 (c) or 11 (d) MCC-I-E^k complexes. **a–d**, Top left, the DIC image of the T-cell-APC conjugate overlaid with the fura-2 fluorescence in false colour (loaded in T cells only) at the time point of the three-dimensional PE fluorescence snapshot. Bottom left, a false-colour reconstruction of the SAV-PE-labelled peptide-MHC fluorescence in the T-cell-APC interface, as viewed from the T cell looking at the APC. Top right, a plot of the time course of the intracellular calcium concentration in the responding T cell, as measured by fura ratioing^{3,13}. Bottom right, a line scan plotting the measured fluorescence signal as a function of position along a line drawn through the interface reconstruction. Arrows on the reconstruction images indicate the location of single peptides; circles indicate clusters where peptides are close enough together to overlap their diffraction-limited spots. The total number of peptides present in each interface is indicated in the margin; where overlap of PE singlet spots occurred, the total number of peptides was determined by dividing the fluorescence intensity of clusters by the mean intensity of single PE spots from the same experiment. Arrows on the calcium time courses indicate the time point at which three-dimensional peptide-MHC fluorescence snapshots were obtained. The 'stretched' appearance of fluorescence in the z direction of interface reconstructions is an artefact of the algorithm used to present a contiguous image (Methods).

20 min (Fig. 2d). A similar trend was observed for 2B4 T-cell blasts responding to SAV-PE-labelled APCs (data not shown).

To quantify the dose response of calcium signalling to defined numbers of peptides, we integrated the fura ratio data (Fig. 2, top right) over 10 min following the initial calcium increase; Fig. 3 shows this integrated signal plotted against the number of specific peptides in the T-cell-APC interface detected early after calcium flux for both 5C.C7 and 2B4 T-cell blasts. For both 5C.C7 and 2B4 T cells, the rapid rise in calcium signalling in response to only a few peptides, which formed a plateau on exposure to 20–30 specific complexes in the interface, emphasizes the high sensitivity of the responding cells. This relationship between ligand number and calcium increase seems to be general, as lines developed from the cytotoxic T cell 2C (ref. 10) give a similar dose-response curve when exposed to defined numbers of an agonist peptide bound to a MHC class I molecule (M. Purbhoo, unpublished data).

T-cell responses to a few agonist peptides were not rare events, as indicated by the frequency of responses (Table 1). T cells stopped and formed close contacts with APCs in 95% of frontal contacts (that is, the leading edge of the T cell contacting peptide) with 1–10 peptides. To characterize this phenomenon further, we prepared streptavidin-coated fluorescent polymer nanospheres (diameter 25 nm) to label biotinylated peptides. These photostable nanospheres could be tracked over the whole time course of an experiment but bound biotinylated peptides inefficiently (~10%) and thus could not be used for quantification. But in examples where the measured calcium signal in a responding T cell was consistent with the number of nanosphere-labelled complexes in the interface (calibrated by measuring the calcium signalling of the same preparation of T cells versus peptide number on SAV-PE-labelled APCs, see above), T cells showed very high efficiencies of antigen detection. In particular, T cells responded by stopping and fluxing calcium in 67% of encounters with single nanosphere-labelled peptides if frontal contact was made (Supplementary Information). But T cells making side or rear contact with antigenic peptide were generally unresponsive to the antigen, consistent with previous reports¹⁴. We conclude that the T cells used here can detect single peptides with a very high degree of efficiency.

The results with the nanospheres also show that T cells do not collect peptides from distal regions of the APC surface; once T cells encounter a ligand, 'scanning' of the APC halts and only those

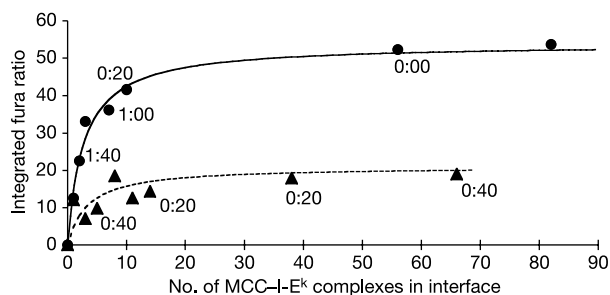


Figure 3 Dose response of T-cell calcium signals to specific peptide-MHC complexes in the T-cell-APC interface. To quantify early calcium signalling in T cells, fura-2 ratios were measured every 20 s in responding T cells and then integrated for 10 min from the time of the initial calcium increase. Integrated fura ratios are shown versus the number of MCC-I-E^k complexes detected by SAV-PE labelling in the T-cell-APC interface for 5C.C7 T-cell blasts (circles) or 2B4 T-cell blasts (triangles). Each point represents data from a single cell. PE snapshots were collected at early time points just after calcium flux; representative times relative to calcium flux for four of the points on each data set are shown. Lines are best fit trends using a simple Langmuir equation. Data were pooled from several experiments carried out on different days with T cells from the same lymph node preparation. Blasts from different lymph node preparations were found to have varying maximum values for the plateau in calcium signalling but similar overall trends.

Table 1 T-cell 'stop' signals and calcium flux in response to defined numbers of peptide-MHC ligands

Condition	No. of agonist pMHC in contact zone*	% of migrating T cells stopping and forming tight interface on ligand contact†	% of stopped cells with Ca ²⁺ signal‡	n
No receptor blocking	0	7.7	100.0	13
	1–10	95.2	100.0	21
	11–25	100.0	100.0	7
	>25	100.0	100.0	7
CD4 blocked	1–10	9.1	0	11
	11–25	25.0	100.0	4
	>25	100.0	87.5	8

Results are tabulated from *n* contacts formed normally or under conditions in which CD4 binding was blocked with antibodies (Methods).

*The number of MCC-I-E^k complexes detected by PE fluorescence in the reconstructed image of the T-cell-APC interface.

†The percentage of *n* T cells responding to contact with a given number of peptide-MHC complexes.

‡The percentage of stopped T cells that also fluxed calcium, as assessed by fura ratiometry.

peptides initially present at the interface are 'seen' by the T cell. Thus, our single time point snapshots of SAV-PE-labelled peptides should be an accurate reflection of the number of peptides initially encountered by the T cell. This conclusion is also supported by the fact that calcium signalling and synapse formation (see below) were similar whether the number of peptides was 'read out' seconds after initial calcium flux or after 10–12 min (data not shown).

Full T-cell activation and subsequent functional activity correlates with a characteristic ordering of cell-surface and signalling molecules at the T-cell-APC interface known as an "immunological synapse"^{1–3}. A hallmark of synapse formation is the accumulation of I-E^k on the APC, which is driven by a complementary accumulation of TCR on the T cell^{1,15}. We found that such a stable, central accumulation of I-E^k forms when ten or more agonist peptides are present (Fig. 4b, c). This is not the result of a general clustering of membrane molecules on the APC in the synapse, because yellow fluorescent protein (YFP)-labelled H-2K^b (MHC class I) molecules expressed by CH27 cells showed no significant accumulation at the interface under conditions that result in robust T-cell stimulation (Fig. 4d). Ten or more peptides at the interface are also sufficient to produce in the T-cell-APC interface a peripheral ring of green fluorescent protein (GFP)-labelled ICAM-1, another indicator of synapse formation (data not shown).

Consistent with earlier results¹⁵, this accumulation of I-E^k corresponds to a very large excess of endogenous MHC molecules relative to the number of specific peptides. For example, the central cores of the synapses shown in Fig. 4b and c both comprise about 6,000 I-E^k

molecules as compared with 10 and 24 agonist peptides, respectively (Supplementary Information). This is about 20% of the total I-E^k-GFP molecules in the area of contact and may represent the fraction of endogenous peptide-I-E^k complexes that is capable of binding to the 5C.C7 TCR. It would be expected that many endogenous peptide-MHCs would not be able to bind a given TCR owing to interference with the bound peptide (as shown for MCC Lys99Glu; ref. 24), but until now there have been no estimates of the fraction of peptide-MHCs that might be compatible.

Previous work has shown that CD4 increases T-cell sensitivity to antigen by 10–100-fold^{16–18}. Consistent with this, we found that blocking CD4 has a marked effect on the response of T cells to APCs loaded with less than about 25 antigenic peptide-MHC complexes. Shown in Fig. 5a are the integrated calcium increases in T cells responding to defined numbers of peptides with or without block-

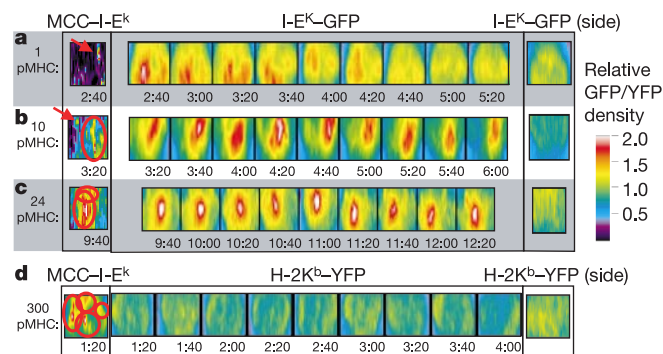


Figure 4 Effect of peptide number on endogenous accumulation of I-E^k. APCs (**a–c**, A20 I-E^k-GFP; **d**, CH27 H-2K^b-YFP) were pulsed with biotinylated MCC peptide, labelled with SAV-PE and mixed with 5C.C7 T cells and imaged. Representative data are shown. Each panel depicts from left to right: the reconstructed false-colour fluorescence of labelled peptide-MHC (pMHC) in the T-cell-APC interface, as viewed from the T cell looking at the APC; nine time-lapse frames of the reconstructed GFP/YFP fluorescence in the T-cell-APC interface (starting immediately after the peptide-MHC snapshot); and, for comparison, the GFP/YFP fluorescence on the APC surface reconstructed for one (noninteracting) side of the APC at the initial GFP time point. Relative GFP accumulations are calibrated by the false-colour scale bar. The times indicated for each image (min:s) are relative to the initial increase in T-cell calcium. Arrows and circles on the peptide-MHC reconstructions indicate peptide locations as in Fig. 2.

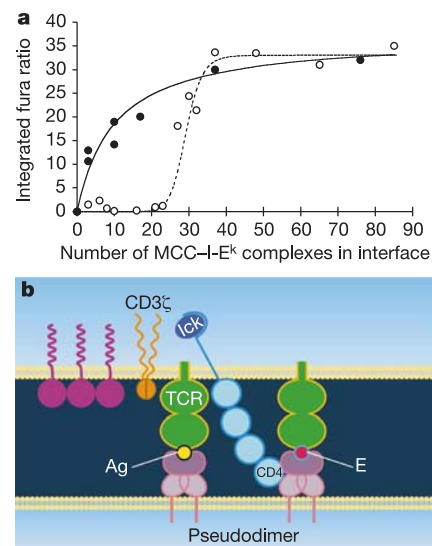


Figure 5 Role of CD4 in the T-cell response to low numbers of agonist peptides. **a**, Effect of blocking CD4 on T-cell calcium signalling. Time-lapse microscopy was used to image 5C.C7 T cells interacting with SAV-PE-labelled A20 APCs expressing I-E^k-GFP with or without CD4 binding blocked by antibodies. Integrated fura ratios are plotted as a measure of early calcium signalling (see Fig. 3) for the case of no receptor blocking (filled circles) or blocked CD4 binding (open circles). Each point represents data from a single cell; data were pooled from several experiments carried out on different days using T cells from the same lymph node preparation. Lines are guides to the eye. **b**, CD4-dependent pseudodimer model for early activation of T cells. When a TCR binds to an agonist peptide-MHC complex (Ag), a conformational change is triggered in an associated CD4 that allows it to bind to nearby MHC molecules. If a nearby MHC has an endogenous peptide (E) that can be bound by the TCR, this could create a 'pseudodimer' structure that provides an initial trigger for TCR signalling. A simple conformational shift in CD4 could also have the effect of bringing the tyrosine kinase Lck into closer proximity with the CD3 ζ chains of its associated TCR, promoting the initial tyrosine phosphorylation that occurs very early in T-cell activation.

ing by an antibody against CD4 (results with an Fab fragment were identical; Supplementary Information). Without blocking, T cells generally showed calcium signalling on contact with even one peptide. By contrast, when CD4 binding was prevented, T cells failed to flux calcium unless 25 or more specific complexes were present in the contact area. Calcium signalling then recovered sharply and attained amounts comparable to unblocked cells at higher peptide densities, with synapse formation proceeding normally. CD4 blockade not only abrogated calcium signalling, but in most cases T cells failed to receive a 'stop' signal in response to low densities of agonist peptide (Table 1). Thus, there is a marked dependence on CD4 for the recognition of low numbers of specific peptides. But these effects can be overcome with a much higher density of ligand, and T-cell stimulation then seems to proceed normally.

Our finding that T cells can detect even a single ligand is reminiscent of experiments showing that retinal cells of the eye can detect single photons¹⁹, which indicates that crucial external and internal sensory systems are geared towards the maximum possible sensitivity. This result is unexpected because even very high concentrations of monomeric peptide–MHC in solution do not trigger CD4-positive T-cell activation. Instead, this requires peptide–MHC trimers²⁰ or dimers²¹. Many other cell-surface receptors also activate cells by dimerization or aggregation²². So how could a single ligand elicit a calcium response and a 'stop' signal? The fact that this acute sensitivity is dependant on CD4 may be relevant in light of the structural work on CD4 class II MHC complexes. In particular, it has been shown that CD4 binds the MHC at almost a 90° degree angle²³, which makes it unlikely that a CD4 molecule that is closely associated with a TCR (as many are in activated T cells^{24,25}) can bind to the same peptide–MHC ligand that the TCR is contacting. Instead, it seems more likely that it binds to an adjacent MHC. If that MHC carries an endogenous peptide that is compatible with TCR binding, then a 'pseudodimer' might result (Fig. 5b) in which a weak interaction between the TCR and the peptide–MHC complex could be stabilized by the weak binding of CD4 to the MHC²⁶. We propose that endogenous ligands do not normally trigger T cells by this combination of TCR and CD4, either because this interaction is too weak to form stable dimers without an agonist ligand or because CD4 is normally unable to bind MHC class II molecules^{26,27}. In the absence of CD4, the number of agonist ligands required to have any effect on T cells went up markedly to about 25. A reason for this might be that, because endogenous ligands can no longer be used in this situation, activation depends on the dimerization of agonist ligands, which could only occur randomly if there were many more present in the cell–cell interface.

Methods

Cells and constructs

We stimulated 5C.C7 and 2B4 primary T cells taken from the lymph nodes of transgenic mice in the presence of 1 μ M MCC (88–103) peptide and maintained them in RPMI 1640 medium plus 10% fetal calf serum (FCS), 2 mM L-glutamine, 50 μ M β -mercaptoethanol and penicillin/streptomycin. T-cell blasts were used on days 4–8 for imaging experiments. Two B-cell lymphoma cell lines, CH27 and A20, were used as APCs. CH27s transfected with full-length ICAM-1 fused to GFP have been described²⁸. We generated CH27 cells stably expressing YFP-tagged H-2K^b (linker sequence AAGGGGSGGGGSGGGGDPVAT) by retroviral transfection of the fusion gene construct using the Phoenix retroviral system²⁹. A20 cells expressing I-E^k fused to the carboxy terminus of the β -chain have been described¹⁵. All APCs were maintained in the same medium as T-cell blasts.

Peptides, antibodies and reagents

We prepared peptide MCC(88–103) by solid-phase synthesis with a biotinylated amino terminus connected to the MHC-binding sequence by a flexible linker (Research Genetics), biotin–SGGGSGGGANERADLIAYLKQATK. Underlined residues were synthesized as D-stereoisomers to avoid possible proteolytic cleavage of residues extending outside the MHC binding groove. The monoclonal antibody D4, which recognizes MCC peptides bound to I-E^k (ref. 7) was conjugated with R-phycoerythrin (PE) using succinimidyl 4-[N-maleimidomethyl]cyclohexane-1-carboxylate, and purified by size-exclusion chromatography on S200 resin (Amersham Pharmacia). We obtained PE-conjugated streptavidin from Pharmingen. Size-exclusion chromatography on high-

resolution Sephacryl S300 resin (Amersham Pharmacia) showed a single monomodal peak, and ultraviolet-visible spectroscopy confirmed a 1:1 ratio of SAV to PE in the conjugate. Streptavidin-conjugated fluorescent nanospheres were prepared by coupling a streptavidin derivative (Neutravidin, Pierce) to the surface of 26-nm diameter red fluorescent carboxylate nanospheres (Molecular Probes). Briefly, 2.25×10^{14} sonicated spheres were added to 2.25 ml of 0.1 M MES buffer, 0.5 M NaCl and 0.05% Tween 20 (pH 6.0). We added EDC (1-ethyl-3-[3-dimethylaminopropyl]carbodiimide hydrochloride) and N-hydroxy succinimide to concentrations of 2 mM and 5 mM, respectively, and rotated the suspension at room temperature in the dark for 15 min. The remaining EDC was quenched by adding mercaptoethanol to 20 mM for 5 min, and 2 mg of streptavidin was added and allowed to react with the activated spheres for 2 h at room temperature in the dark. The reaction was stopped by adding glycine to 100 mM for 30 min, followed by dialysis for 24 h through a membrane with a 300,000 molecular mass cutoff into PBS containing 0.05% Tween 20. The streptavidin-coupled nanospheres were purified by size-exclusion chromatography on S200 resin (Amersham Pharmacia).

Peptide labelling and loading determination

APCs were pulsed with biotinylated peptide for 1 h at 37 °C in 7% CO₂ in medium at 4×10^6 cells ml⁻¹, washed twice with 10 ml of PBS, and resuspended in 50 μ l of 1% bovine serum albumin in PBS (Pierce) with 0.5% sodium azide (to prevent internalization or export of MHC) and either 1–10 μ g ml⁻¹ SAV-PE or 10¹¹ ml⁻¹ SAV fluorescent nanospheres. Cells were incubated at room temperature in the dark for 20 min, washed with 10 ml of PBS, and placed on ice until used.

We characterized peptide labelling by SAV-PE or D4-PE by a FACSscan flow cytometer (Becton Dickinson). APCs were pulsed with peptide and labelled with SAV-PE or D4-PE at a range of concentrations. We quantified binding of the PE-labelled molecules by calibrating the measured fluorescence per PE molecule using Quantibrite PE calibration beads (Pharmingen). Total peptide–MHC complexes present for different peptide pulse concentrations were determined by Scatchard analysis of binding data. In preliminary experiments, nonspecific binding of the biotinylated MCC peptide to cells lacking I-E^k MHC molecules was undetectable (as assessed by negligible staining by D4-PE or SAV-PE on biotinylated MCC-pulsed C57Bl6 splenocytes (I-A^b); data not shown).

Further confirmation of FACS labelling and calibration was obtained by labelling 0.1 or 1 μ M peptide-pulsed cells with a saturating amount of SAV-PE and measuring the fluorescence per cell by three-dimensional digital microscopy. We converted the microscopic fluorescence determined for 50 cells to total peptide number by calibrating against the fluorescence from discrete single PE molecules observed in the same samples.

Microscopy and image analysis

Microscopy and data analysis was carried out on a Zeiss/Universal Imaging three-dimensional time-lapse system using a $\times 63$ objective as described³. Images were obtained with 0.319 μ m per pixel in x and y and 1- μ m steps in the z direction (perpendicular to the substrate). Before imaging, T cells were loaded with 10 μ M fura-2-AM (Molecular Probes) in medium for 25 min at 37 °C. To examine the effects of CD4 receptor blockade, blocking antibodies were added to the appropriate cells at 20 μ g ml⁻¹ for 20 min at 4 °C before imaging. (The binding of antibodies to Fc receptors on APCs did not seem to be a factor because identical results were obtained with or without pre-blocking Fc receptors using a combination of antibodies against CD16 and CD32 and excess irrelevant mouse IgG.) Fura-loaded T cells and APCs labelled as above were mixed together in minimal imaging medium (PBS plus 5% FCS, 0.5 mM CaCl₂ and 0.1 mM MgCl₂) in Labtek eight-well chambered coverslips (Nunc) and imaged on a stage heated to 37 °C. Fura-2 ratios, as a measure of intracellular calcium concentration, were calculated from fluorescence images as described^{3,13,15,30}. In addition to fura-2 fluorescence imaging, the distribution of GFP-tagged cell-surface molecules (I-E^k or ICAM-1) on the APC was also tracked to provide a second indicator of a productive T-cell response and synapse formation^{3,28}. Although the large fluorescence yield from PE permitted detection of single peptide–MHC complexes, the rapid photobleaching of the label limited three-dimensional observations to a single time point during T-cell–APC interactions. GFP-labelled protein distributions could be tracked for longer periods of time, but only subsequent to the PE 'snapshot', owing to the overlap in the excitation bands of these fluorochromes. Thus, time-lapse experiments were carried out in the following manner: first, T cells were mixed with peptide-pulsed, labelled APCs; second, a time-lapse recording of two-dimensional differential interference contrast (DIC) and fura-2 fluorescence images was initiated; third, after some T-cell–APC conjugates had formed (as assessed by live monitoring of DIC/fura images), a three-dimensional snapshot of the peptide–MHC distribution was recorded; last, time-lapse recording of GFP-labelled protein three-dimensional distributions at each interval (in addition to the DIC and fura-2 images) was begun immediately after the PE snapshot. PE fluorescence was excited by a single-band filter centred at 555 nm and collected through a single-band emission filter at 580–650 nm; each two-dimensional fluorescence plane was collected with a 400-ms exposure. GFP/YFP fluorescence images were obtained by excitation at 480/510 nm and emission at 510/535 nm, respectively, with an exposure time of 300 ms. We collected time-lapse images at intervals of 20 s.

We determined the number of specific peptide–MHC complexes present in T-cell–APC contact zones by reconstructing three-dimensional images of the interface. Interface views were built using maximum intensity projections. To present a contiguous image, the data in the z axis (collected in 1- μ m intervals) were interpolated to obtain a pixel size in z equivalent to that in the x and y directions (0.319 μ m per pixel). Single PE molecules were found to have a narrow distribution of total integrated intensities (and maximum intensities); thus, the intensity of single molecules could be calibrated for each experiment from discrete, well-defined single PE spots. For interfaces where clustering of molecules made the total number of PEs present unclear by simple observation, determinations were

made by dividing total cluster intensities by the mean intensity of single molecules. Images shown for all conditions are representative of 3–10 (or more) experiments conducted on different days. □

Received 6 June; accepted 5 August 2002; doi:10.1038/nature01076.

- Monks, C. R. E., Freiberg, B. A., Kupfer, H., Sciaki, N. & Kupfer, A. Three-dimensional segregation of supramolecular activation clusters in T cells. *Nature* **395**, 82–86 (1998).
- Grakoui, A. *et al.* The immunological synapse: a molecular machine controlling T cell activation. *Science* **285**, 221–227 (1999).
- Krummel, M. F., Sjaastad, M. D., Wulfig, C. & Davis, M. M. Differential clustering of CD4 and CD3 ζ during T cell recognition. *Science* **289**, 1349–1352 (2000).
- Harding, C. V. & Unanue, E. R. Quantitation of antigen-presenting cell MHC class II/peptide complexes necessary for T-cell stimulation. *Nature* **346**, 574–576 (1990).
- Demotz, S., Grey, H. M. & Sette, A. The minimal number of class-I MHC antigen complexes needed for T-cell activation. *Science* **249**, 1028–1030 (1990).
- Kimachi, K., Croft, M. & Grey, H. M. The minimal number of antigen-major histocompatibility complex class II complexes required for activation of naive and primed T cells. *Eur. J. Immunol.* **27**, 3310–3317 (1997).
- Reay, P. A. *et al.* Determination of the relationship between T cell responsiveness and the number of MHC-peptide complexes using specific monoclonal antibodies. *J. Immunol.* **164**, 5626–5634 (2000).
- Christinck, E. R., Lüscher, M. A., Barber, B. H. & Williams, D. B. Peptide binding to class I MHC on living cells and quantitation of complexes required for CTL lysis. *Nature* **352**, 67–70 (1991).
- Brower, R. C. *et al.* Minimal requirements for peptide mediated activation of CD8⁺ CTL. *Mol. Immunol.* **31**, 1285–1293 (1994).
- Sykulev, Y., Joo, M., Vturina, I., Tsomides, T. J. & Eisen, H. N. Evidence that a single peptide–MHC complex on a target cell can elicit a cytolytic T cell response. *Immunity* **4**, 565–571 (1996).
- Triantafyllou, K., Triantafyllou, M. & Wilson, K. M. Phycobiliprotein–Fab conjugates as probes for single particle fluorescence imaging. *Cytometry* **41**, 226–234 (2000).
- Sako, Y., Minoguchi, S. & Yanagida, T. Single-molecule imaging of EGFR signalling on the surface of living cells. *Nature Cell Biol.* **2**, 168–172 (2000).
- Tsien, R. Y. Fluorescent probes of cell signaling. *Annu. Rev. Neurosci.* **12**, 227–253 (1989).
- Negulescu, P. A., Krasieva, T. B., Khan, A., Kerschbaum, H. H. & Cahalan, M. D. Polarity of T cell shape, motility, and sensitivity to antigen. *Immunity* **4**, 421–430 (1996).
- Wulfig, C. *et al.* Costimulation and endogenous MHC ligands contribute to T cell recognition. *Nature Immunol.* **3**, 42–47 (2002).
- Marrack, P. *et al.* The major histocompatibility complex-restricted antigen receptor on T cells. II. Role of the L3T4 product. *J. Exp. Med.* **158**, 1077–1091 (1983).
- Hampel, J., Chien, Y. H. & Davis, M. M. CD4 augments the response of a T cell to agonist but not to antagonist ligands. *Immunity* **7**, 379–385 (1997).
- Vidal, K., Daniel, C., Hill, M., Littman, D. R. & Allen, P. M. Differential requirements for CD4 in TCR–ligand interactions. *J. Immunol.* **163**, 4811–4818 (1999).
- Baylor, D. A., Lamb, T. D. & Yau, K. W. Responses of retinal rods to single photons. *J. Physiol. (Lond.)* **288**, 613–634 (1979).
- Boniface, J. J. *et al.* Initiation of signal transduction through the T cell receptor requires the peptide multivalent engagement of MHC ligands. *Immunity* **9**, 459–466 (1998).
- Cochran, J. R., Cameron, T. O. & Stern, L. J. The relationship of MHC–peptide binding and T cell activation probed using chemically defined MHC class II oligomers. *Immunity* **12**, 241–250 (2000).
- Heldin, C. H. Dimerization of cell surface receptors in signal transduction. *Cell* **80**, 213–223 (1995).
- Wang, J. H. *et al.* Crystal structure of the human CD4 N-terminal two-domain fragment complexed to a class II MHC molecule. *Proc. Natl Acad. Sci. USA* **98**, 10799–10804 (2001).
- Anderson, P., Blue, M. L. & Schlossman, S. F. Comodulation of CD3 and CD4. Evidence for a specific association between CD4 and approximately 5% of the CD3/T cell receptor complexes on helper T lymphocytes. *J. Immunol.* **140**, 1732–1737 (1988).
- Beyers, A. D., Spruyt, L. L. & Williams, A. F. Molecular associations between the T-lymphocyte antigen receptor complex and the surface antigens CD2, CD4, or CD8 and CD5. *Proc. Natl Acad. Sci. USA* **89**, 2945–2949 (1992).
- Xiong, Y., Kern, P., Chang, H.-C. & Reinherz, E. L. T cell receptor binding to a pMHCII ligand is kinetically distinct from and independent of CD4. *J. Biol. Chem.* **276**, 5659–5667 (2001).
- Crawford, F., Kozono, H., White, J., Marrack, P. & Kappler, J. Detection of antigen-specific T cells with multivalent soluble class II MHC covalent peptide complexes. *Immunity* **8**, 675–682 (1998).
- Wulfig, C., Sjaastad, M. D. & Davis, M. M. Visualizing the dynamics of T cell activation: Intracellular adhesion molecule 1 migrates rapidly to the T cell/B cell interface and acts to sustain calcium levels. *Proc. Natl Acad. Sci. USA* **95**, 6302–6307 (1998).
- Pear, W. S., Nolan, G. P., Scott, M. L. & Baltimore, D. Production of high-titer helper-free retroviruses by transient transfection. *Proc. Natl Acad. Sci. USA* **90**, 8392–8396 (1993).
- Wulfig, C. & Davis, M. M. A receptor/cytoskeletal movement triggered by costimulation during T cell activation. *Science* **282**, 2266–2269 (1998).

Supplementary Information accompanies the paper on Nature's website (<http://www.nature.com/nature>).

Acknowledgements We thank B. Newsome, W. E. Moerner, J. Huppa, L. Wu and P. Ebert for discussions. This work was supported by grants (to M.M.D.) from the NIH and the Howard Hughes Medical Institute. D.J.I. was supported by the Cancer Research Fund of the Damon Runyon-Walter Winchell Foundation Fellowship; M.A.P. was supported by the Howard Hughes Medical Institute; and M.K. was supported by a fellowship from the Alfred Benzon Foundation.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to M.M.D. (e-mail: mdavis@cmgm.stanford.edu).

The prolyl isomerase Pin1 is a regulator of p53 in genotoxic response

Hongwu Zheng[†], Han You[†], Xiao Zhen Zhou[†], Stephen A. Murray[†], Takafumi Uchida[‡], Gerburg Wulfig[†], Ling Gu^{*}, Xiaoren Tang[§], Kun Ping Lu[†] & Zhi-Xiong Jim Xiao^{*§}

^{*} Department of Biochemistry and [§] Department of Medicine, Boston University School of Medicine, Boston, Massachusetts 02118, USA

[†] Department of Medicine, Beth Israel Deaconess Medical Center, Harvard Medical School, Boston, Massachusetts 02115, USA

[‡] Department of Pathology, Institute of Development, Aging and Cancer, Tohoku University, Sendai, Japan

p53 is activated in response to various genotoxic stresses resulting in cell cycle arrest or apoptosis^{1,2}. It is well documented that DNA damage leads to phosphorylation and activation of p53 (refs 1–3), yet how p53 is activated is still not fully understood. Here we report that DNA damage specifically induces p53 phosphorylation on Ser/Thr-Pro motifs, which facilitates its interaction with Pin1, a member of peptidyl-prolyl isomerase^{4–9}. Furthermore, the interaction of Pin1 with p53 is dependent on the phosphorylation that is induced by DNA damage. Consequently, Pin1 stimulates the DNA-binding activity and trans-activation function of p53. The Pin1-mediated p53 activation requires the WW domain, a phosphorylated Ser/Thr-Pro motif interaction module, and the isomerase activity of Pin1. Moreover,

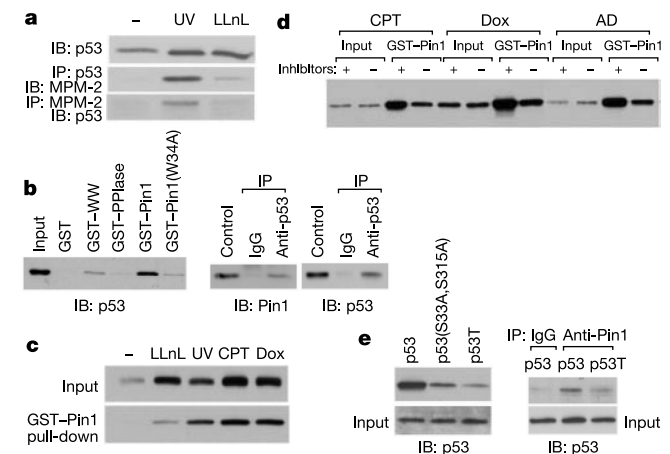


Figure 1 DNA damage induces p53 phosphorylation on Ser-Pro motifs and p53–Pin1 interaction. **a**, Cell lysates from ultraviolet (UV)-irradiated (25 J m^{−2}) or LLnL-treated (50 μM) U2-OS cells were immunoprecipitated (IP) using a p53 antibody (FL393) or antibody MPM-2, and subjected to immunoblotting analysis (IB) using MPM-2 or p53 antibody. **b**, p53–Pin1 interaction *in vitro* and *in vivo*. Doxorubicin-treated U2-OS cell lysates were used for GST–Pin1 pull-down assay and western blotted for p53. Cell lysates from doxorubicin-treated MCF-7 cells were immunoprecipitated with an antibody for p53 (Pab1801), Pin1, or control IgG, which were subjected to western blot for Pin1 or p53, respectively. **c**, Cell lysates from U2-OS cells treated with various DNA-damaging agents or LLnL were subjected to GST–Pin1 pull-down assay and western blotted for p53. CPT, camptothecin; Dox, doxorubicin. **d**, Cell lysates were incubated with calf intestine phosphatase in the presence (+) or absence (−) of phosphatase inhibitors before GST–Pin1 pull-down assay. AD, actinomycin D. **e**, Saos-2 cells were transfected with wild-type p53, the double mutant p53(S33A,S315A) or the p53T triple mutant p53(S33A,S315A,T81A), followed by ultraviolet (40 J m^{−2}) treatment. Cell lysates were subjected to GST–Pin1 pull-down assay (left panel) or to immunoprecipitation with Pin1 antibody or IgG followed by western blot for p53.

Pin1-deficient cells are defective in p53 activation and timely accumulation of p53 protein, and exhibit an impaired checkpoint control in response to DNA damage. Together, these data suggest a mechanism for p53 regulation in cellular response to genotoxic stress.

The protein phosphorylation on Ser/Thr-Pro motifs is a principal regulatory mechanism in cellular response to a variety of signals^{5,10,11}. Thus, we examined phosphorylation of p53 on DNA damage using a monoclonal antibody MPM-2 that specifically recognizes the pSer/Thr-Pro motif^{7,12}. Ultraviolet irradiation not only led to the accumulation of p53 protein but also a greatly enhanced phosphorylation on Ser/Thr-Pro motifs in contrast to the treatment with the proteasome inhibitor N-acetyl-Leu-Leu-norleucinal (LLnL) (Fig. 1a). Similarly, treatment of cells with doxorubicin or actinomycin D also triggered p53 phosphorylation on Ser/Thr-Pro motif(s) (data not shown).

Many MPM-2 antigens can also interact with the peptidyl-prolyl isomerase Pin1 (ref. 13), which specifically binds to pSer/Thr-Pro motifs and isomerizes the pSer/Thr-Pro peptide bond. Pin1 consists of a WW domain (a protein-protein interaction domain consisting of 35–40 amino acids with two conserved tryptophans), which is involved in substrate interaction, and the isomerase domain (called PPIase domain). When tested, p53 interacted with full-length Pin1

and with the WW domain, but not with the PPIase domain alone (Fig. 1b). A single point mutation in the WW domain (Pin1(W34A)), a binding module for pSer/Thr-Pro^{7,14}, lost most of its binding activity to p53. In addition, an *in vivo* p53–Pin1 protein complex was readily detected upon DNA damage (Fig. 1b). Furthermore, DNA-damage-induced phosphorylation of p53 enhanced its interaction with Pin1, as cells treated with ultraviolet irradiation, camptothecin (CPT) or doxorubicin exhibited a higher p53-binding activity to Pin1 than LLnL-treated cells (Fig. 1c), and removal of p53 phospho-groups by phosphatase treatment led to a clearly diminished p53–Pin1 interaction (Fig. 1d).

There are four Ser-Pro (Ser 33, Ser 46, Ser 127 and Ser 315) and two Thr-Pro (Thr 81 and Thr 150) motifs on human p53 protein. Single mutations on these Ser-Pro or Thr-Pro sites did not lead to marked reduction of the p53–Pin1 interaction (Supplementary Fig. 1b). However, a double point mutant, p53(S33A,S315A), showed less binding to Pin1, and the p53T triple point mutant p53(S33A,S315A,T81A) exhibited further reduced binding activity for Pin1 both *in vitro* and *in vivo* (Fig. 1e), suggesting that these three Ser/Thr-Pro sites are important for the p53–Pin1 interaction. Notably, phosphorylation on Ser 33, Ser 315 and Thr 81 can be induced on DNA damage by p38 stress kinase, CDC2 kinase or Jun N-terminal kinase, respectively^{15–20}.

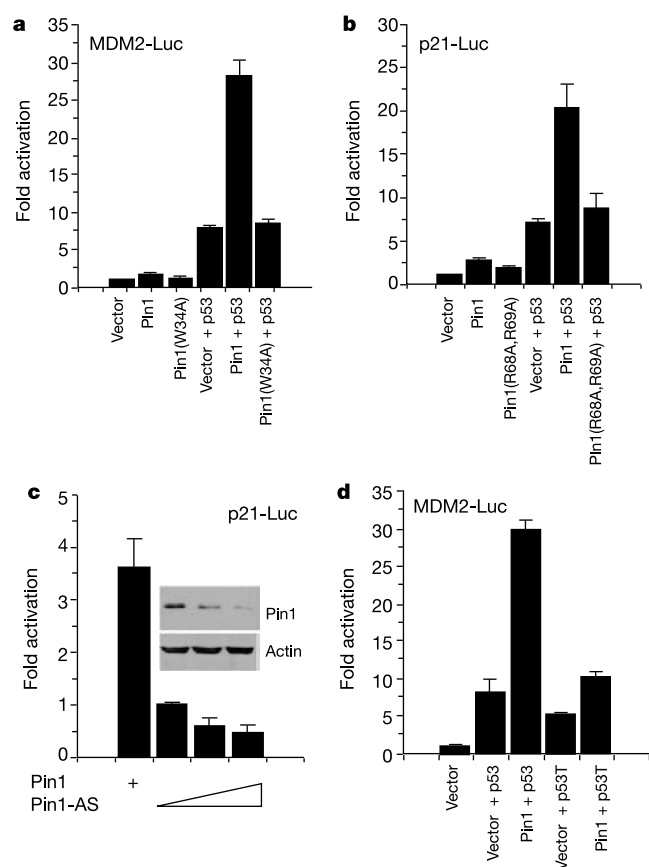


Figure 2 Pin1 enhances p53 transcription activity. **a**, U2-OS cells were co-transfected with a mixture of plasmids expressing p53, β -galactosidase, Pin1, or a Pin1 mutant, in the presence of MDM2-luciferase (**a**) or p21-luciferase (**b**) reporter construct. The luciferase activity was normalized to β -galactosidase activity and presented as fold activation. **c**, Anti-sense Pin1 attenuates p53 transactivation activity. U2-OS cells were co-transfected with p21-luciferase reporter and an expression plasmid encoding anti-sense Pin (Pin1-AS). The endogenous Pin1 protein level on Pin1-AS expression is shown (inset). **d**, Pin1 stimulates transactivation activity of wild-type p53 but not mutant p53T defective in Pin1 interaction. U2-OS cells were co-transfected with MDM2-luciferase reporter and Pin1 in the presence of p53 or p53T.

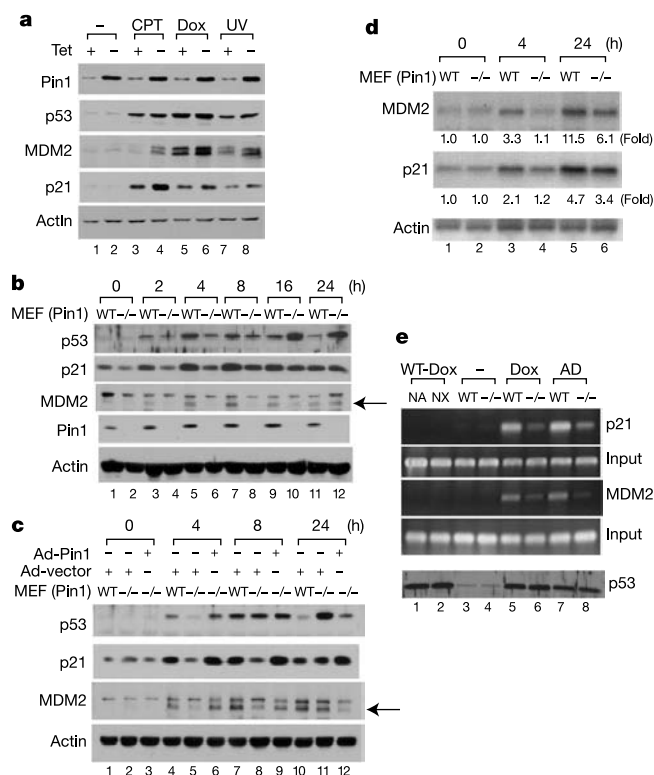


Figure 3 Pin1 activates p53 in response to DNA damage. **a**, Pin1-inducible, stable U2-OS cells were treated with CPT (2 μ M), doxorubicin (0.4 μ g ml⁻¹) or ultraviolet irradiation (UV) (25 J m⁻²) in the presence (+) or absence (–) of tetracycline (Tet). Cell lysates were subjected to western blot analysis, as indicated. **b**, MEF cells were treated with doxorubicin (0.4 μ g ml⁻¹), collected at the indicated time point and subjected to western blot analysis. WT, wild type. **c**, MEF cells were infected with either control adenovirus or adenovirus-Pin1 and treated with doxorubicin (0.4 μ g ml⁻¹) before lysis at the indicated time points. **d**, Northern blot analysis of wild-type MEFs or *Pin1*^{-/-} MEFs treated with doxorubicin for 4 or 24 h. The mRNA expression level was normalized to actin and presented as fold increase. **e**, Chromatin immunoprecipitation analysis on p53 DNA-binding activity in MEFs untreated or treated with either doxorubicin (0.4 μ g ml⁻¹) for 8 h or actinomycin D (5 nM) for 12 h. NA, normal mouse IgG; NX, non-crosslink. Western blot analysis for p53 protein is shown.

Next, we examined the effect of Pin1 on the transactivation function of p53. Both MDM2- and p21-luciferase reporter activities were markedly increased when p53 was co-transfected with wild-type Pin1, but not with mutant Pin1 defective in the p53 interaction (Pin1(W34A)) or defective in isomerase activity (Pin1(R68A,R69A))^{7,14} (Fig. 2a, b). Expression of a Pin1 antisense construct²¹ led to reduced p21-luciferase reporter activity (Fig. 2c), presumably due to the reduction of endogenous Pin1 protein expression (Fig. 2c, inset). Furthermore, Pin1 had little effect on p53 reporter activity in *p53*^{-/-} Saos-2 cells (Supplementary Fig. 1c), suggesting that Pin1 stimulation on the MDM2- or p21-luciferase reporters is dependent on p53. Consistently, the p53T mutant exhibited a significantly reduced response to Pin1 (Fig. 2d).

We also investigated the effect of Pin1 on activation of p53 upon DNA damage. Induction of Pin1 expression from inducible U2-OS cells by removal of tetracycline resulted in a clear increase in both p21 and MDM2 protein expression when treated with DNA-damaging agents (Fig. 3a). Similar observations were obtained by either transient transfection of wild-type Pin1 or infection of a recombinant Pin1 adenovirus (Ad-Pin1), followed by treatment with either actinomycin D or CPT (Supplementary Fig. 2).

As ectopic expression of Pin1 can activate the transactivation activity of p53, we examined the impact of Pin1 deletion on p53 activity. DNA-damage-mediated upregulation of p53 transactivation activity was impaired in *Pin1*^{-/-} cells (Fig. 3b, c). For example, 8 h after doxorubicin treatment, p53 transactivation activity was reduced in *Pin1*^{-/-} cells compared with wild-type mouse embryo fibroblast cells (MEFs), as indicated by comparable levels of p53 and a clearly reduced level of expression of p21 and MDM2 proteins (Fig. 3b, c, lanes 7 and 8). Furthermore, despite significantly higher protein levels of p53 in *Pin1*^{-/-} cells compared

with wild-type MEFs at 24 h after treatment, there was no apparent increase in p21 and MDM2 protein expression (Fig. 3b, lanes 11 and 12, and c, lanes 10 and 11). The impaired p53 transactivation activity in *Pin1*^{-/-} MEFs on treatment with doxorubicin was restored by infection with the Pin1-expressing recombinant adenovirus (Fig. 3c). Consistently, on DNA damage, the messenger RNA levels of both p21 and MDM2 were lower in *Pin1*^{-/-} cells compared with wild-type MEFs (Fig. 3d). Moreover, the impaired p53 transactivation activity in *Pin1*^{-/-} cells could be due to the reduced p53 DNA-binding activity as indicated by chromatin immunoprecipitation (chIP) assay²² (Fig. 3e) and electrophoretic mobility shift assay (EMSA) (Supplementary Fig. 3b, lanes 3 and 4).

Loss of Pin1 also seems to impact on the stability of p53 in MEFs on DNA damage. At early stages of cellular response to doxorubicin treatment, accumulation of p53 protein was impaired in *Pin1*^{-/-} cells (Fig. 3b, c). Similar results were obtained at 16 h after ultraviolet irradiation (Supplementary Fig. 2d). These data suggest that Pin1 could be involved in the timely accumulation of p53 on DNA damage. Whereas p53 protein levels accumulated and then started to decrease 16 h after doxorubicin treatment—as expected—in wild-type MEFs (Fig. 3b), in *Pin1*^{-/-} cells p53 protein levels remained markedly high even 24 h after doxorubicin treatment (Fig. 3b, c). As Pin1 is important for p53 transactivation activity, it is possible that this reduced p53 protein turnover in *Pin1*^{-/-} cells is due to impaired upregulation of MDM2 protein, which functions as the E3 ligase for p53. Normal protein expression of p53, p21 and MDM2 was restored in *Pin1*^{-/-} MEFs by infection with Pin1 adenovirus (Fig. 3c). Taken together, these data indicate that loss of Pin1 leads to defective activation of p53 and a defect in the timely accumulation of p53 protein on DNA damage.

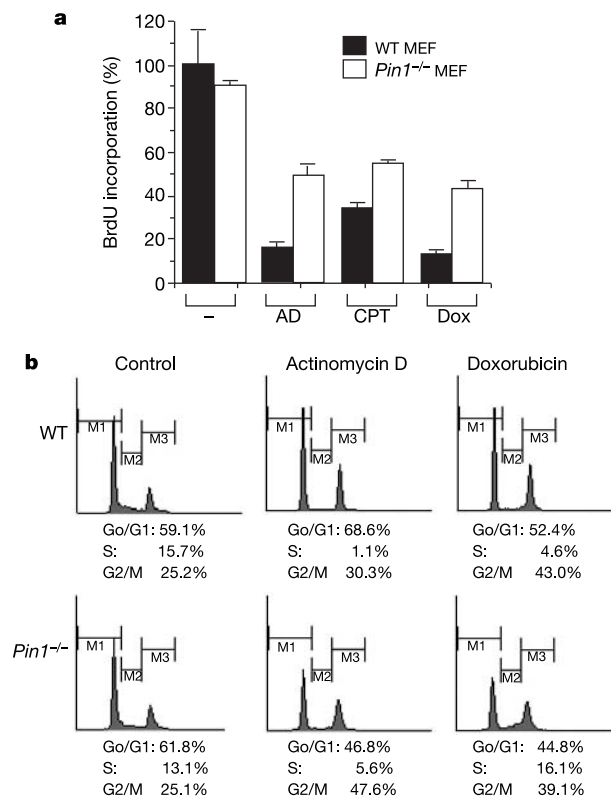


Figure 4 Cell cycle arrest induced by DNA damage is impaired in *Pin1*^{-/-} MEF cells. MEF cells were treated with actinomycin D (AD, 5 nM), camptothecin (CPT, 2 μ M), or doxorubicin (Dox, 0.4 μ g ml⁻¹). **a**, Sixteen hours after treatment, cells were incubated with BrdU (10 μ M) for a further 3 h. Cells were fixed with methanol and stained with

fluorescein isothiocyanate (FITC)-conjugated anti-BrdU-antibody and counter-stained with DAPI. **b**, Cells were collected after 16 h of treatment and subjected to propidium iodide staining and FACS analysis.

As the activity of p53 is altered in the absence of Pin1, we investigated whether the cell cycle is also affected in Pin1-deficient MEFs on DNA damage. As shown in Fig. 4a, under normal growth conditions the number of 5-bromodeoxyuridine (BrdU) positive cells was slightly lower in *Pin1*^{-/-} MEFs. However, on DNA damage, the percentage of BrdU positive cells was markedly higher in *Pin1*^{-/-} cells compared with wild-type MEFs. Furthermore, fluorescence-activated cell sorting (FACS) analysis showed a greater population of cells at the S phase in *Pin1*^{-/-} MEFs on DNA damage (Fig. 4b), presumably due to the impaired p53 activity in the absence of Pin1. These data suggest that Pin1 is important in p53-mediated checkpoint control in response to genotoxic stresses.

The impact of Pin1 on p53 upon DNA damage is demonstrated by the finding that Pin1 binds to phosphorylated p53 and is required for p53 activity and the timely activation of p53. Pin1 functions as a phosphorylation-dependent isomerase^{7,9,14}, which requires both its WW and isomerase domains^{6,14}. Notably, the activities of both of these Pin1 domains are required for its effects on p53. Thus, Pin1 may somehow alter the conformation of pSer/Thr-Pro motifs in p53 to modulate its activities and biological functions. Further work, however, is needed to elucidate the exact function of Pin1 in the regulation of p53.

Pin1 is involved in the regulation of various cellular processes^{4,5,23–25}. Here, we show that it is involved in p53-mediated checkpoint control on DNA damage. Notably, Pin1 has been implicated in human tumorigenesis. It has been demonstrated that Pin1 regulates both the stability and localization of β -catenin²⁶. Pin1 can also cooperate with Ras signalling in the regulation of the JNK and c-Jun pathways to modulate cyclin D1 expression²⁷. In addition, overproduction of Pin1 is reported in breast cancer²⁷, in close correlation with the role of Pin1 in the activation of cyclin D1 (ref. 28). As loss of Pin1 attenuates apoptosis induced by DNA damage in E1A transformed MEFs (Supplementary Fig. 4), it is possible that Pin1 is important in the activation of the p53-mediated apoptotic pathway in transformed cells on DNA damage. Thus, aberrant expression of Pin1, in combination with certain oncogenic signals such as Ras or β -catenin, may provide a stress signal in triggering p53 activation, which provides a safeguard against oncogenic proliferation and imposes a selective pressure for functional inactivation of p53 during the course of tumour progression. Further work is needed to investigate the role of Pin1 in human tumorigenesis. □

Methods

Antibodies, plasmids and recombinant adenoviruses

Actinomycin D, camptothecin, doxorubicin and LLNL were from Sigma. The p53 antibodies (DO-1, 1801, FL393) and actin (C-11) were from Santa Cruz Biotechnology. Antibodies specific for p21^{Waf1/Cip1} (BD Pharmingen), MPM-2 (Upstate Biotechnology) or haemagglutinin (BabCO) were obtained commercially. Monoclonal antibodies for MDM2 and polyclonal antibody against Pin1 were described^{13,29}. The full-length human Pin1 complementary DNA was subcloned into vectors including pTRE (Clontech) and pCDNA-3 (Invitrogen) with or without an haemagglutinin epitope tag. p53 and Pin1 mutants were generated by site-directed mutagenesis using the QuikChange site-directed mutagenesis kit (Stratagene) and confirmed by DNA sequencing. The Pin1 recombinant adenovirus was generated by polymerase chain reaction (PCR) amplification of the haemagglutinin-tagged human Pin1 and cloned into the vector pAdTrack-cytomegalovirus (a gift from B. Vogelstein)³⁰.

Culture, transfection, reporter assays and ChIP analysis

Pin1 tetracycline-inducible U2-OS cell lines were established by using the Tet-Off gene expression system (Clontech). For reporter assays, Saos-2 cells and U2-OS cells were transfected either by the calcium-phosphate precipitation technique or with FuGENE-6 (Roche Diagnostics). Cells were lysed using reporter lysis buffer (Promega) and processed for measuring luciferase activity, which was normalized to β -gal activity. Chromatin immunoprecipitation analysis was performed as described previously²². In brief, adherent MEFs untreated or treated with either doxorubicin (0.4 μ g ml⁻¹) or actinomycin D (5 nM) were crosslinked to chromatin with 1% formaldehyde for 10 min at 37°C. p53-DNA complexes were immunoprecipitated from nuclear extracts by using antibodies against mouse p53 (Ab-1 and Ab-4; Oncogene Research Products), recovered with protein A/G beads and treated with proteinase K to purify DNA. PCR was performed by amplifying fragments of the promoter sequence. We used the following primers: p21, 5' end,

5'-CCTTCTATCAGCCCCAGGATACC-3', 3' end, 5'-GGGACGTCCTTAATTATCTGGGGTC-3'; MDM2, 5' end, 5'-GGTGCCTGGTCCCGACTGCCGGG-3', 3' end, 5'-CCGAGAGGGTCCCCAGGGGTGTCC-3'.

GST pull-down and co-immunoprecipitation

Cells were lysed in 50 mM Tris-HCl (pH 7.5), 200 mM NaCl, 50 mM NaF, 1 mM sodium orthovanadate, 10% glycerol, 1% NP-40, 1 mM dithiothreitol (DTT) and proteinase inhibitors. For *in vitro* phosphatase treatment, cell lysates were incubated with calf intestine phosphatase (0.2 U μ l⁻¹) at 30°C for 45 min in the presence or absence of the phosphatase inhibitors (50 mM NaF and 1 mM sodium orthovanadate). Glutathione S-transferase (GST)-Pin1 pull-down experiments were performed by incubating cell lysates for 2 h at 4°C with 20 μ l agarose beads loaded with various GST-Pin1 proteins or control GST^{13,29}. The co-immunoprecipitation experiments were performed as described previously²⁵.

Proliferation assay and FACS analysis

The cell proliferation assay was performed using the BrdU-labelling kit (Roche Diagnostics). Cells presenting normal nuclear 4,6-diamidino-2-phenylindole (DAPI) staining were counted first and were scored for BrdU staining. At least 100 cells were counted for each point from three independent experiments. The fraction of BrdU-labelled cells in the non-treated wild-type MEF cell population was given an arbitrary volume of 100%. For FACS analysis, cells were collected, re-suspended in 1 ml solution containing 0.1% Triton X-100, 50 μ g ml⁻¹ RNase A and 50 μ g ml⁻¹ propidium iodide. At least 10,000 cells were collected at each time point by FACScan and analysed with the CellQuest program (Becton Dickinson).

Received 28 June; accepted 12 September 2002; doi:10.1038/nature01116.

Published online 2 October 2002.

- Lakin, N. D. & Jackson, S. P. Regulation of p53 in response to DNA damage. *Oncogene* **18**, 7644–7655 (1999).
- Woods, D. B. & Vousden, K. H. Regulation of p53 function. *Exp. Cell Res.* **264**, 56–66 (2001).
- Meek, D. W. Mechanisms of switching on p53: a role for covalent modification. *Oncogene* **18**, 7666–7675 (1999).
- Hunter, T. Prolyl isomerases and nuclear function. *Cell* **92**, 141–143 (1998).
- Lu, K. P. Phosphorylation-dependent prolyl isomerization: a novel cell cycle regulatory mechanism. *Prog. Cell Cycle Res.* **4**, 83–96 (2000).
- Lu, K. P., Hanes, S. D. & Hunter, T. A human peptidyl-prolyl isomerase essential for regulation of mitosis. *Nature* **380**, 544–547 (1996).
- Yaffe, M. B. *et al.* Sequence-specific and phosphorylation-dependent proline isomerization: a potential mitotic regulatory mechanism. *Science* **278**, 1957–1960 (1997).
- Ranganathan, R., Lu, K. P., Hunter, T. & Noel, J. P. Structural and functional analysis of the mitotic rotamase Pin1 suggests substrate recognition is phosphorylation dependent. *Cell* **89**, 875–886 (1997).
- Stukenberg, P. T. & Kirschner, M. W. Pin1 acts catalytically to promote a conformational change in cdc25. *Mol. Cell* **7**, 1071–1083 (2001).
- Nigg, E. A. Cyclin-dependent protein kinases: key regulators of the eukaryotic cell cycle. *Bioessays* **17**, 471–480 (1995).
- Sudol, M. & Hunter, T. New wrinkles for an old domain. *Cell* **103**, 1001–1004 (2000).
- Davis, F. M., Tsao, T. Y., Fowler, S. K. & Rao, P. N. Monoclonal antibodies to mitotic cells. *Proc. Natl Acad. Sci. USA* **80**, 2926–2930 (1983).
- Shen, M., Stukenberg, P. T., Kirschner, M. W. & Lu, K. P. The essential mitotic peptidyl-prolyl isomerase Pin1 binds to and regulates mitosis-specific phosphoproteins. *Genes Dev.* **12**, 706–720 (1998).
- Lu, P. J., Zhou, X. Z., Shen, M. & Lu, K. P. Function of WW domains as phosphoserine- or phosphothreonine-binding modules. *Science* **283**, 1325–1328 (1999).
- Bulavin, D. V. *et al.* Phosphorylation of human p53 by p38 kinase coordinates N-terminal phosphorylation and apoptosis in response to UV radiation. *EMBO J.* **18**, 6845–6854 (1999).
- Sanchez-Prieto, R., Rojas, J. M., Taya, Y. & Gutkind, J. S. A role for the p38 mitogen-activated protein kinase pathway in the transcriptional activation of p53 on genotoxic stress by chemotherapeutic agents. *Cancer Res.* **60**, 2464–2472 (2000).
- Buschmann, T. *et al.* Jun NH2-terminal kinase phosphorylation of p53 on Thr-81 is important for p53 stabilization and transcriptional activities in response to stress. *Mol. Cell Biol.* **21**, 2743–2754 (2001).
- Milner, J., Cook, A. & Mason, J. p53 is associated with p34 cdc2 in transformed cells. *EMBO J.* **9**, 2885–2889 (1990).
- Wang, Y. & Prives, C. Increased and altered DNA binding of human p53 by S and G2/M but not G1 cyclin-dependent kinases. *Nature* **376**, 88–91 (1995).
- Blaydes, J. P. *et al.* Stoichiometric phosphorylation of human p53 at Ser315 stimulates p53-dependent transcription. *J. Biol. Chem.* **276**, 4699–4708 (2001).
- Rippmann, J. F. *et al.* Phosphorylation-dependent proline isomerization catalyzed by Pin1 is essential for tumour cell survival and entry into mitosis. *Cell Growth Differ.* **11**, 409–416 (2000).
- Flores, E. R. *et al.* p63 and p73 are required for p53-dependent apoptosis in response to DNA damage. *Nature* **416**, 560–564 (2002).
- Hsu, T., McRackan, D., Vincent, T. S. & Gert De Couet, H. *Drosophila* Pin1 prolyl isomerase Dodo is a MAP kinase signal responder during oogenesis. *Nature Cell Biol.* **3**, 538–543 (2001).
- Winkler, K. E., Swenson, K. I., Kornbluth, S. & Means, A. R. Requirement of the prolyl isomerase Pin1 for the replication checkpoint. *Science* **287**, 1644–1647 (2000).
- You, H. *et al.* IGF-1 induces Pin1 expression in promoting cell cycle S-phase entry. *J. Cell Biochem.* **84**, 211–216 (2002).
- Ryo, A., Nakamura, M., Wulf, G. G., Liou, Y.-C. & Lu, K. L. Prolyl isomerase Pin1 regulates turnover and subcellular localization of β -catenin by inhibiting its interaction with APC. *Nature Cell Biol.* **3**, 793–801 (2001).
- Wulf, G. M. *et al.* Pin1 is overexpressed in breast cancer and cooperates with Ras signaling in increasing the transcriptional activity of c-Jun towards cyclin D1. *EMBO J.* **20**, 3459–3472 (2001).
- Liou, Y. C. *et al.* Loss of Pin1 function in the mouse causes phenotypes resembling cyclin D1-null phenotypes. *Proc. Natl Acad. Sci. USA* **22**, 1335–1340 (2002).
- Xiao, Z. X. *et al.* Interaction between the retinoblastoma protein and the oncoprotein MDM2. *Nature* **375**, 694–698 (1995).

30. He, T. C. *et al.* A simplified system for generating recombinant adenoviruses. *Proc. Natl Acad. Sci. USA* 95, 2509–2514 (1998).

Supplementary Information accompanies the paper on *Nature's* website (<http://www.nature.com/nature>).

Acknowledgements We thank B. Vogelstein, K. Vousden and T. Jacks for plasmids; J. Chen for 2A-10 antibody; and G. Del Sal for discussion and sharing unpublished data. We also thank E. R. Flores for E1A-retrovirus and advice on ChIP analysis, and Y. Zhang for technical assistance on cell cycle analysis. This work was supported by the NIH (Z.-X.X.) and Department of Defense (Z.-X.X.).

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to Z.-X.X. (e-mail: jxiao@bu.edu).

The prolyl isomerase Pin1 reveals a mechanism to control p53 functions after genotoxic insults

Paola Zacchi^{*†‡}, Monica Gostissa^{*‡}, Takafumi Uchida[§], Clio Salvagno^{*†}, Fabio Avolio^{*}, Stefano Volinia^{||}, Ze'ev Ronai[¶], Giovanni Blandino[#], Claudio Schneider^{*☆} & Giannino Del Sal^{*†}

^{*} Laboratorio Nazionale CIB, AREA Science Park, Padriciano 99, 34012 Trieste, Italy

[†] Dipartimento di Biochimica, Biofisica e Chimica delle Macromolecole, Università degli Studi di Trieste, via L. Giorgeri 1, 34100, Trieste, Italy

[§] Department of Pathology, Institute of Development, Aging and Cancer, Tohoku University, Sendai 980-8575, Japan

^{||} Università di Ferrara, Sezione di Istologia ed Embriologia, Dipartimento di Morfologia ed Embriologia, via Fossato di Mortara 64/b, 44100, Ferrara, Italy

[¶] The Rutenberg Cancer Center, Mount Sinai School of Medicine, One Gustave L. Levy Place, Box 1130, New York 10029-6574, USA

[#] Molecular Oncogenesis Laboratory, Regina Elena Cancer Institute, via Messi d'oro 156, 00158 Rome, Italy

[☆] Dipartimento di Scienze e Tecnologie Biomediche, Università degli Studi di Udine, p. le Kolbe 1, 33100 Udine, Italy

[‡] These authors contributed equally to this work

The tumour suppressor p53 is important in the cell decision to either arrest cell cycle progression or induce apoptosis in response to a variety of stimuli. p53 post-translational modifications and association with other proteins have been implicated in the regulation of its stability and transcriptional activities^{1,2}. Here we report that, on DNA damage, p53 interacts with Pin1, a peptidyl-prolyl isomerase³, which regulates the function of many proteins involved in cell cycle control and apoptosis^{4–6}. The interaction is strictly dependent on p53 phosphorylation, and requires Ser 33, Thr 81 and Ser 315. On binding, Pin1 generates conformational changes in p53, enhancing its transactivation activity. Stabilization of p53 is impaired in UV-treated *Pin1*^{−/−} cells owing to its inability to efficiently dissociate from Mdm2. As a consequence, a reduced p53-dependent response was detected in *Pin1*^{−/−} cells, and this correlates with a diminished transcriptional activation of some p53-regulated genes. Our results suggest that, following stress-induced phosphorylation, p53 needs to form a complex with Pin1 and to undergo a conformational change to fulfil its biological roles.

The peptidyl-prolyl isomerase Pin1 has recently emerged as an important regulator of cell proliferation and DNA-replication checkpoint^{3,7}. Pin1 interacts with a number of phosphoproteins through recognition of pSer/Thr-Pro motifs by its amino-terminal

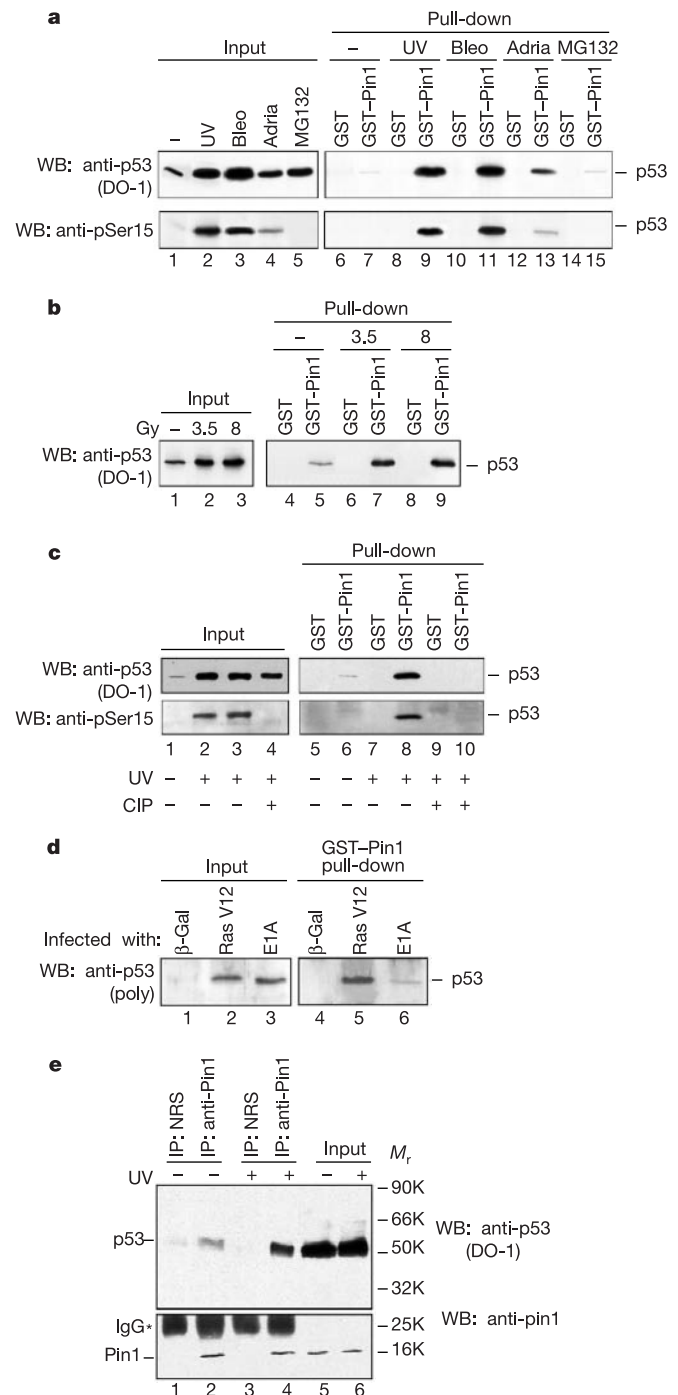


Figure 1 Pin1 interacts with activated p53 in a phosphorylation-dependent manner. **a**, Lysates from U2-OS cells treated with different DNA-damaging agents or with proteasome inhibitor MG132 were subjected to GST or GST-Pin1 pull-down followed by immunoblotting with DO-1 antibody. Anti-pSer15 antibody was used to detect DNA-damage-induced phosphorylation on p53. WB, western blot; UV, ultraviolet radiation; Bleo, bleomycin; Adria, adriamycin. **b**, U2-OS cells were γ -irradiated and processed as in **a**. **c**, Lysates from UV-irradiated U2-OS cells were treated with calf intestinal phosphatase (CIP) before incubation with GST and GST-Pin1. **d**, NIH3T3 cells infected with retroviruses expressing β -galactosidase (β -Gal), activated Ras (RasV12) or E1A were analysed as in **a**. Murine p53 was detected by anti-p53 polyclonal antibody. **e**, Lysates from 293 cells, UV irradiated and untreated, were immunoprecipitated with anti-Pin1 antibody or with normal rabbit serum (NRS). Immunoprecipitates were analysed by western blotting with DO-1 and anti-Pin1 antibodies. IgG, immunoglobulin- γ . Asterisk indicates the light chain of the antibodies used for immunoprecipitation.

WW domain^{8,9}. By promoting the *cis-trans* isomerization of this peptide bond through its carboxy-terminal prolyl isomerase (PPIase)³ domain, it regulates the function of its substrates.

Analysis of the amino-acid sequence of human p53 revealed several Pin1 consensus motifs, which are phosphorylated by members of the MAP kinase¹⁰ and the cell-cycle-dependent kinase¹¹ (Cdk) families on genotoxic insults. Therefore we tested whether Pin1 could associate with p53 by using a glutathione-S-transferase (GST)-Pin1 pull-down assay⁶. Pin1 bound p53 with high affinity only on treating the cells with several types of DNA-damaging agents, including ultraviolet radiation (UV), ionizing radiation and chemotherapeutic drugs (Fig. 1a, b). This binding is not simply due to quantitative changes in p53 levels, as no interaction with GST-Pin1 was detected when p53 was accumulated by treatment with proteasome inhibitor (Fig. 1a, lane 15). The interaction relied entirely on p53 phosphorylation, as demonstrated by the lack of binding when p53 was dephosphorylated by phosphatase treatment (Fig. 1c), and required a functional Pin1-WW domain (Supplementary Fig. S1a). Similar results were obtained with other human and murine cell lines (not shown). In addition to DNA damage, overexpression of oncogenes activating p53^{12,13} promoted its interaction with GST-Pin1 (Fig. 1d).

The *in vivo* association between Pin1 and p53 was detected by coimmunoprecipitation between either overexpressed (Supplementary Fig. S1b) or endogenous proteins (Fig. 1e), and the binding was greatly enhanced on UV treatment.

To identify the Pin1 binding sites on p53, we generated the

corresponding Ser/Thr to Ala point mutants, and assayed their ability to bind GST-Pin1 on overexpression in SaOS-2 cells and UV treatment. Under these conditions, wild-type p53 (p53WT) and p53(S46A) exhibited a similar affinity for Pin1, whereas p53(S33A), p53(T81A) and p53(S315A) showed a reduced binding. This reduction was more pronounced in the case of the double and triple mutants (Fig. 2a), thus suggesting that these three sites are all critical for Pin1 binding.

Next we verified whether the prolyl-isomerase activity of Pin1 induces conformational changes in p53 by performing a partial proteolysis assay¹⁴ on *in vitro*-translated p53WT, which is phosphorylated on Pin1 consensus, as judged by immunoprecipitation with the pSer/Thr-Pro specific MPM-2 antibody⁶, and binds to Pin1 (not shown). His-tagged p53WT was incubated with GST or GST-Pin1 before addition of subtilisin. Pin1 protected p53 from proteolytic cleavage, whereas lack of protection was observed on p53 being stripped of its phosphates by treatment with calf intestinal phosphatase (CIP; Fig. 2b, left panel). Importantly this effect was not due to steric hindrance exerted by Pin1 binding, but was dependent only on its isomerase activity (Fig. 2b, right panel), because GST-Pin1(C109A), a mutant that is catalytically impaired but fully competent to bind its substrates⁷, was completely ineffective.

Changes in p53 conformation may impinge on its transactivation functions. We therefore performed reporter assays in SaOS-2 cells overexpressing p53WT. The ectopic expression of Pin1, but not Pin1-WW mutant, exerted a stimulatory effect toward Bax- and

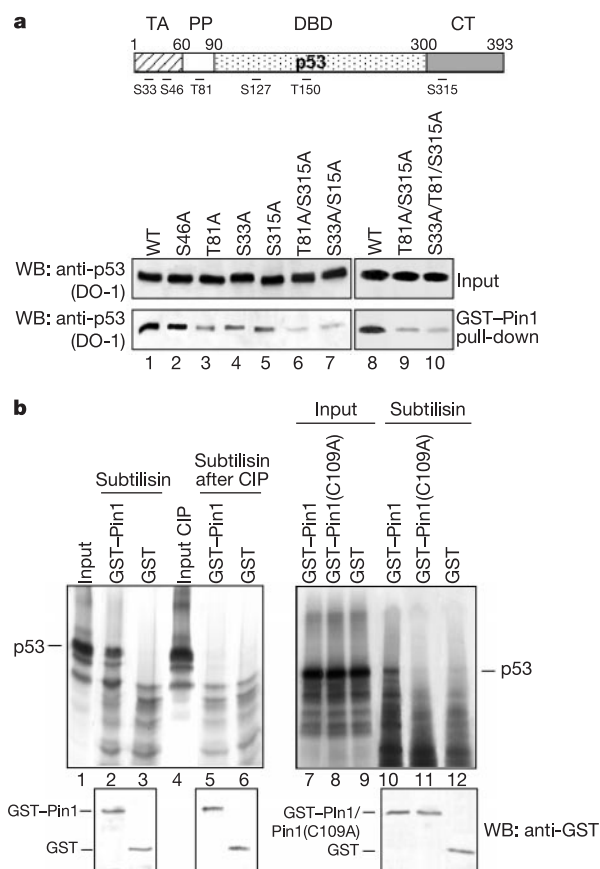


Figure 2 Phosphorylated Ser 33, Thr 81 and Ser 315 are required for the interaction with Pin1, which drives conformational changes in p53. **a**, Different Ser/Thr-Ala p53 point mutants were assayed for interaction with GST-Pin1 as described in Fig. 1 on overexpression in SaOS-2 cells and UV treatment. A representation of p53 domains is presented above (TA, transactivation domain; PP, polyproline region; DBD, DNA binding domain; CT, C-terminal domain), together with the positions of putative Pin1 consensus

motifs. WT, wild type. **b**, Subtilisin cleavage was performed on His-tagged p53 after incubation with GST, GST-Pin1 or GST-Pin1(C109A) (right panel), and on p53 dephosphorylation (left panel). The reaction products were analysed by SDS-PAGE and autoradiography. The amount of the various GST-fusions were verified by western blotting (lower panels).

pG13-luciferase reporters. This effect was not observed when Pin1 was co-expressed with p53 mutants S33A/S315A, T81A/S315A or S33A/T81A/S315A, thus indicating that it is a consequence of a direct interaction between Pin1 and p53 (Supplementary Fig. S2).

As the data obtained so far indicated Pin1 to be an important mediator of p53 activation, we investigated whether it would influence p53 accumulation in *Pin1*^{-/-} mouse embryo fibroblasts (MEFs)¹⁵. *Pin1*^{+/+} and *Pin1*^{-/-} MEFs, which express comparable levels of p53 under unstressed conditions, were exposed to UV light and checked for p53 induction 12 h later. Whereas a marked increase of p53 levels was detectable in *Pin1*^{+/+} cells, p53 induction was reduced in *Pin1*-deficient MEFs (Fig. 3a, upper panel), though the protein was still phosphorylated at Ser15 (Fig. 3a, central panel) indicating that it sensed the insult in both cases. In contrast, p53 was equally well stabilized by proteasome inhibitors (Fig. 3a, lanes 3 and 6). The reduced p53 protein levels were not due to changes in p53 transcription as judged by northern blot analysis (not shown). Notably, the impairment in p53 stabilization was fully rescued when *Pin1*^{-/-} MEFs were infected with a retrovirus expressing Pin1, but not Pin1(C109A), further demonstrating the requirement of Pin1 isomerase activity for p53 stabilization (Fig. 3b). The lower steady-state levels of p53 in UV-treated *Pin1*^{-/-} cells correlated with a reduced half-life of the protein (Fig. 3c), while no differences were observed in cells untreated or treated with proteasome inhibitors (not shown). Moreover, also in this case, reintroduction of Pin1 was able to efficiently restore a normal p53 half-life (not shown).

Mdm2 is important as a regulatory partner of p53, and disruption of the p53-Mdm2 complex is a key event for p53 stabilization^{1,2}. Therefore, to assess a possible mechanism underlying the impaired p53 stabilization in the absence of Pin1, we analysed the ability of p53 to bind to Mdm2 following DNA damage by

performing pull-down assays (Fig. 3d, lanes 1–6) and coimmunoprecipitation (Fig. 3d, lanes 7–12) assays on lysates normalized for p53 content. This revealed that in UV-treated *Pin1*^{-/-} cells a higher amount of p53 bound Mdm2, whereas no differences between both cell lines were observed when p53 was stabilized by proteasome inhibitors (not shown). These results therefore indicate that conformational changes driven by Pin1 in p53 may also influence its interaction with Mdm2.

Finally we investigated whether the reduced accumulation of p53 in *Pin1*^{-/-} MEFs was affecting p53-mediated apoptotic response. MEFs were UV-treated, fixed 12 or 24 h later and nuclear morphology was analysed by epifluorescence. In *Pin1*^{-/-} MEFs, the number of apoptotic nuclei was reduced by about 50% as compared to the wild-type counterpart (Fig. 4a and Supplementary Fig. S3b), and this reduction was confirmed by fluorescence-activated cell sorting (FACS) analysis (Fig. 4b). In agreement with these observations, northern blot analysis indicated that, under these conditions, two p53-induced pro-apoptotic genes like *Bax* and *Killer/DR5* were not efficiently upregulated in *Pin1*^{-/-} cells (Fig. 4c). To rule out the dependence of the observed effect on the specific type of DNA damage, we performed similar experiments on adriamycin treatment. The apoptotic response was reduced in *Pin1*^{-/-} MEFs, and the pro-apoptotic gene *Killer/DR5* was not induced (Supplementary Fig. S4) while p21 and Mdm2 upregulation was only slightly impaired. Moreover, to evaluate the ability of these cells to arrest cell cycle progression, 5-bromodeoxyuridine (BrdU) incorporation assays were performed following UV irradiation. Whereas at early time points both cells displayed similar arrest, at later time points *Pin1*^{-/-} MEFs showed a higher S-phase re-entry as compared to the wild-type MEFs (Supplementary Fig. S3a).

A striking impairment in the apoptotic response was also observed following UV irradiation in *Pin1*^{-/-} MEFs overexpressing

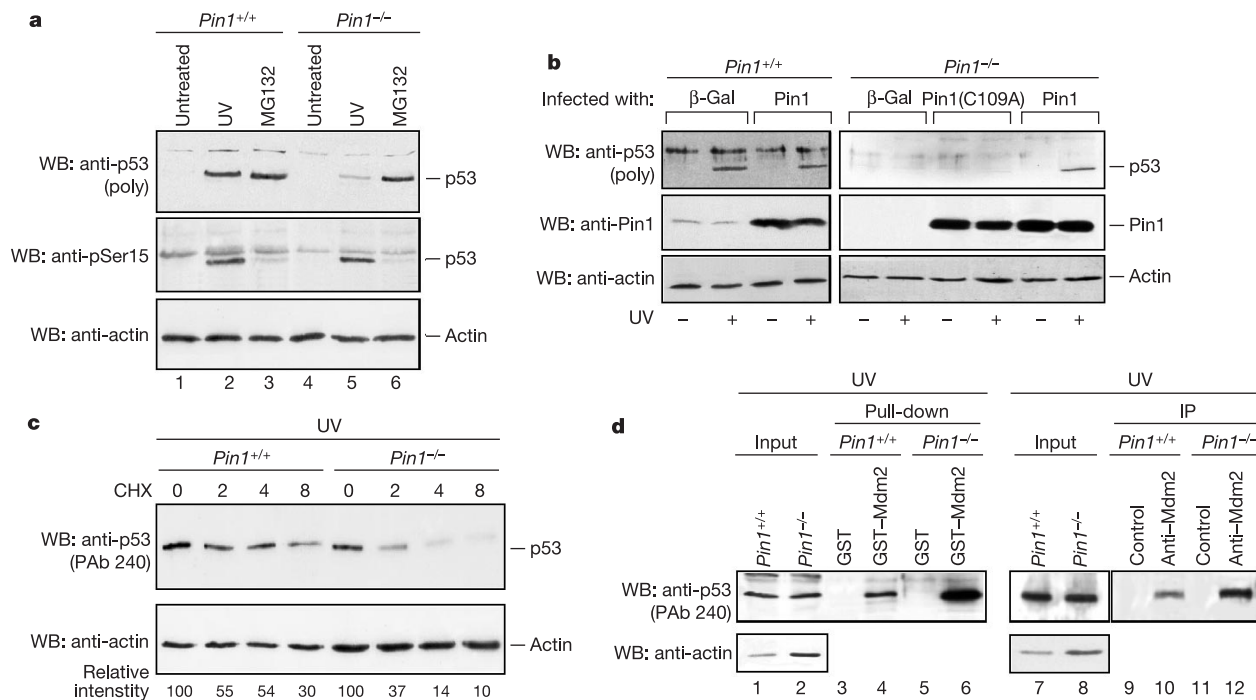


Figure 3 p53 stabilization following DNA damage is reduced in *Pin1*^{-/-} MEFs. **a**, Lysates from *Pin1*^{+/+} and *Pin1*^{-/-} MEFs, treated with UV or MG132, were analysed by immunoblotting with anti-p53 or anti-pSer15 antibodies. **b**, Western blot analysis with anti-p53 or with anti-Pin1 antibodies was performed on lysates from *Pin1*^{+/+} and *Pin1*^{-/-} MEFs infected with retroviruses expressing β-Gal, Pin1 or Pin1(C109A). Cells were either untreated or UV-damaged as indicated. **c**, Lysates from UV-treated *Pin1*^{+/+} and *Pin1*^{-/-} MEFs were collected at the indicated time points following addition of

cycloheximide (CHX), and analysed by immunoblotting with anti-p53 monoclonal antibody. The relative intensity of each p53 band quantified by densitometry is shown below. **d**, Wild-type or *Pin1*^{-/-} MEFs were UV-treated, and lysates were subjected to GST-Mdm2 pull-down (lanes 1–6) or immunoprecipitated with anti-Mdm2 2A10 antibody or with 9E10 as a negative control (lanes 7–12). Mdm2–p53 interaction was analysed by western blotting with Pab240. The amounts of lysate were normalized for p53 content. Anti-actin staining was used as loading control in all the experiments.

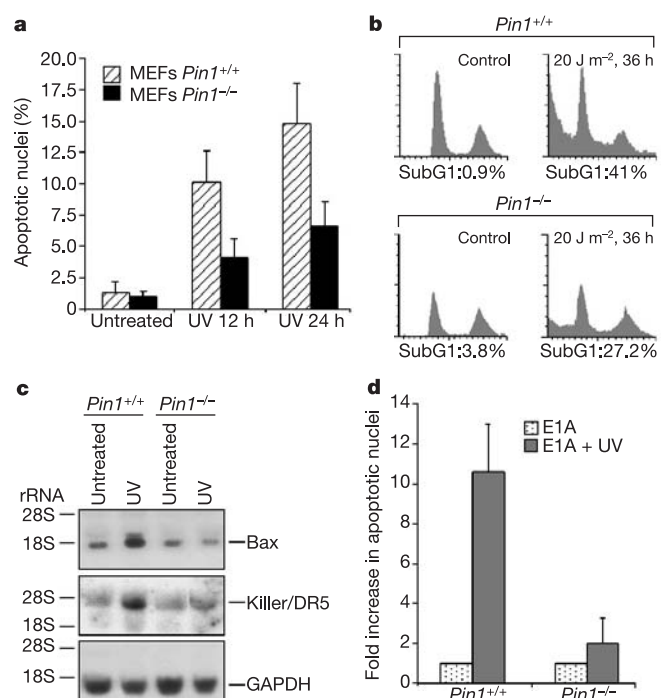


Figure 4 p53 response following DNA damage is impaired in *Pin1*^{-/-} MEFs. **a**, *Pin1*^{+/+} and *Pin1*^{-/-} MEFs were UV-irradiated, and 12 or 24 h later nuclear morphology was assessed by Hoechst staining. Graphs represent the mean of six independent experiments. At least 300 cells were counted in each experiment. **b**, Cells (floating and attached) were collected, and DNA content determined by FACS analysis after staining with propidium iodide. At least 10,000 cells were examined in each experiment. SubG1 refers to apoptotic nuclei. **c**, Total RNA extracted from *Pin1*^{+/+} and *Pin1*^{-/-} MEFs 12 h after UV irradiation was analysed by northern blot with the indicated probes. Blots were probed with GAPDH as a loading control. **d**, *Pin1*^{+/+} and *Pin1*^{-/-} MEFs were infected with retrovirus expressing E1A. Apoptosis was scored as in **a** 12 h after UV treatment.

E1A (Fig. 4d) and Myc (not shown), known to sensitize cells to apoptosis through a p53-dependent mechanism¹⁶, further demonstrating the involvement of p53 in this effect. No apoptosis was indeed observed in *p53*^{-/-} MEFs similarly treated (not shown).

The data presented here allow us to put forward a model in which kinases activated by several stresses target p53, thus promoting the binding of Pin1. This, in turn, generates a conformational change in the protein, which may facilitate additional modifications and influence its ability to complex with other factors, resulting in its full stabilization and functional activation. But p53 regulation depends on the integration of many damage-sensing pathways¹⁷, with the final outcome being influenced by the type and the intensity of the insult, as well as by cell/tissue type¹⁸. In this context, we could not detect significant differences in p53 accumulation and apoptosis in γ -irradiated thymocytes from *Pin1*^{+/+} and *Pin1*^{-/-} mice (not shown). Another level of complexity is introduced by the pleiotropic effects that Pin1, by acting on several substrates, may have on the overall cellular environment.

Pin1 was identified as an essential mitotic regulator⁶, but accumulating evidence has now assigned it an important role in several other processes. It has been shown to regulate c-Jun following JNK (c-Jun N-terminal kinase) phosphorylation, and to modulate cyclin D1 expression¹⁹. Moreover, Pin1 contributes to the regulation of the turnover of β -catenin²⁰, known to affect the levels of cyclin D1²¹ as well as of p53²². Under certain conditions, both cyclin D1 and β -catenin can also induce apoptosis^{23,24}. This evidence, together with our results indicating that p53 is another relevant target of Pin1, identify Pin1 as a critical node of integration

among DNA-damage-induced checkpoint pathways, and further underlines the complexity and interdependence of the mechanisms governing p53 functions.

Methods

Cell lines and DNA damage treatments

U2-OS are human osteosarcoma cells wild-type for p53. MCF-7 is a human mammary breast carcinoma cell line containing wild-type p53. 293 is a human embryonic kidney cell line. *Pin1*^{+/+} and *Pin1*^{-/-} mouse embryo fibroblasts (MEFs) have been described previously¹⁶. For UV damage, cells were subjected to 20 J m⁻² UV-C irradiation in the calibrated area within the tissue culture hood. Bleomycin and adriamycin were added to the culture medium for 12 or 24 h at the concentration of 50 U l⁻¹ and 0.4 μ g ml⁻¹, respectively. For proteasome inhibitor treatment, the following drugs were added to the culture medium for 6 h: MG132 (50 μ M, Sigma), E64D (50 μ M, Sigma) and MG101 (50 μ M, Sigma).

Antibodies

The following primary antibodies were used: FL-393 (rabbit polyclonal anti-p53, SantaCruz); DO-1 (monoclonal anti-p53, SantaCruz); PAb240 (monoclonal anti-p53, SantaCruz); rabbit polyclonal anti-Pin1 raised against human Pin1 expressed in bacteria as a GST fusion protein and affinity-purified by standard procedures; rabbit polyclonal anti-actin (Sigma); 2A10 (monoclonal anti-Mdm2); 9E10 (monoclonal anti-Myc); goat polyclonal anti-GST (Amersham Biosciences); rabbit polyclonal anti-phosphorylated Ser15 (pSer15) antibody (Cell Signaling Biotechnology); rabbit polyclonal anti-phospho Thr81 (pThr81)²⁵.

Plasmids

pcDNA3p53WT has been previously described²⁶. Complementary DNAs encoding the various p53 point mutants were generated by polymerase chain reaction (PCR)-directed mutagenesis, and cloned into pcDNA3. cDNA encoding Pin1 was provided by the I.M.A.G.E. consortium (RZPD, Deutschen HumanGenomeProjekt). pGEX4T1 plasmids containing Pin1, the point mutants Pin1(Y23A) and Pin1(C109A), as well as the isolated WW (GST–Pin1 WW) and isomerase (GST–Pin1 PPIase) domains, were constructed by PCR. All PCR-amplified products were fully sequenced to exclude the possibility of second site mutations. pcDNA3Pin1, pcDNA3Pin1WW, pBABEPin1 and pBABEPin1(C109A) were constructed by subcloning the corresponding cDNAs from pGEX4T1. pBABE-8Gal, pLPC-RasV12 and pBABE-E1A were provided by R. Maestro, M. Serrano and B. Amati, respectively. The p53 luciferase reporters pG13-Luc and Bax-Luc, as well as pGEX-Mdm2, were a gift from M. Oren.

In vitro binding, immunoprecipitation, western blot and CHX treatment

GST pull-down assays from cell lysates were performed as previously described⁶. The lysis buffer was supplemented with phosphatase (1 mM sodium orthovanadate, 5 mM NaF) and protease (phenylmethylsulphonyl fluoride (PMSF) 1 mM and chymostatin, leupeptin, antipain, pepstatin 10 μ g ml⁻¹ each) inhibitors. In the case of phosphatase treatment, 20 U ml⁻¹ of CIP (New England Biolabs) were added to the diluted extracts and the reaction was continued for 30 min at 30 °C.

For p53 and Pin1 immunoprecipitation analysis, cells were collected in PBS pH 8.3 buffer (PBS pH 8.3, EDTA 10 mM, Tween20 0.1%) with inhibitors as above, and lysed by passing through a 26G needle. The p53 and Mdm2 coimmunoprecipitation experiments were performed as previously described²⁷.

For cycloheximide (CHX) treatment, cells were UV-irradiated (20 J m⁻²), then 12 h later the medium was supplemented with 50 μ g ml⁻¹ of CHX; cells were then collected at the indicated time points.

Subtilisin proteolysis

Subtilisin digestion was performed as previously described¹⁵. Briefly, ³⁵S-labelled His-p53 was purified on a nickel column, and incubated with 100 ng of either GST–Pin1WT, GST–Pin1(C109A) or GST in the following buffer: 50 mM HEPES, pH 7.5, 100 mM NaCl, 1 mM MgCl₂, 1 mM dithiothreitol, supplemented with phosphatase inhibitors. After 10 min incubation at 20 °C, subtilisin (6.25 ng μ l⁻¹, Sigma) was added for a further 6 min at 5 °C. The reaction was stopped by the addition of boiling sample buffer containing 8 M urea, and the proteolytic fragments were resolved by 15% SDS–polyacrylamide gel electrophoresis (SDS–PAGE) and visualized by autoradiography.

Retroviral infection

Phoenix packaging cells were transfected with the indicated vectors by a standard calcium phosphate method. After 48 h incubation at 32 °C, the supernatants containing viral particles were collected, and infection was performed as described²⁸. In the case of NIH3T3, cells were further processed for pull-down 3 d after infection. In the case of *Pin1*^{+/+} and *Pin1*^{-/-} MEFs, after one week of selection in puromycin at the final concentration of 2 μ g ml⁻¹, the polyclonal populations of infected cells were analysed for Pin1 expression, expanded and further processed.

Reporter assays and northern blotting

For reporter assays, SaOS-2 cells were transfected with 1 μ g of Luc reporters, 50 ng of p53 and 1 μ g of Pin1 expression plasmids. In all the samples, 50 ng of the reporter pRL-cytomegalovirus (Promega) was included for normalization of the transfection efficiency. 24 h after transfections, cells were lysed and assayed for luciferase activity using the Dual Luciferase kit (Promega).

For northern blotting, 7 µg of RNA prepared by a standard method was loaded in each lane. Hybridizations were performed under high stringency conditions, and exposure times varied between overnight and 3 d at –80 °C with intensifying screens.

Received 24 June; accepted 12 September 2002; doi:10.1038/nature01120.
Published online 2 October 2002.

1. Sionov, R. & Haupt, Y. The cellular response to p53: the decision between life and death. *Oncogene* **18**, 6145–6157 (1999).
2. Vousden, K. H. p53: death star. *Cell* **103**, 691–694 (2000).
3. Lu, K. P., Hanes, S. D. & Hunter, T. A human peptidyl-prolyl isomerase essential for regulation of mitosis. *Nature* **380**, 544–547 (1996).
4. Sudol, M. & Hunter, T. NeW Wrinkles for an old domain. *Cell* **103**, 1001–1004 (2000).
5. Pathan, N., Aime-Sempe, C., Kitada, S., Haldar, S. & Reed, J. C. Microtubule targeting drugs induce Bcl-2 phosphorylation and association with Pin1. *Neoplasia* **3**, 70–79 (2001).
6. Shen, M., Stukenberg, P. T., Kirschner, M. W. & Lu, K. P. The essential mitotic peptidyl-prolyl isomerase Pin1 binds and regulates mitosis-specific phosphoproteins. *Genes Dev.* **12**, 706–720 (1998).
7. Winkler, K. E., Swenson, K. I., Kornbluth, S. & Means, A. R. Requirement of the prolyl isomerase Pin1 for the replication checkpoint. *Science* **287**, 1644–1647 (2000).
8. Lu, P. J., Zhou, X. Z., Shen, M. & Lu, K. P. Function of WW domains as phosphoserine- or phosphothreonine-binding modules. *Science* **283**, 1325–1328 (1999).
9. Verdecia, M. A., Bowman, M. E., Lu, K. P., Hunter, T. & Noel, J. P. Structural basis for phosphoserine-proline recognition by group IV WW domains. *Nature Struct. Biol.* **7**, 639–643 (2000).
10. Bulavin, D. V. *et al.* Phosphorylation of human p53 by p38 kinase coordinates N-terminal phosphorylation and apoptosis in response to UV radiation. *EMBO J.* **18**, 6845–6854 (1999).
11. Wang, Y. & Prives, C. Increased and altered DNA binding of human p53 by S and G2/M but not G1 cyclin-dependent kinases. *Nature* **376**, 88–91 (1995).
12. Serrano, M., Lin, A. W., McCurrach, M. E., Beach, D. & Lowe, S. W. Oncogenic ras provokes premature cell senescence associated with accumulation of p53 and p16INK4a. *Cell* **88**, 593–602 (1997).
13. de Stanchina, E. *et al.* E1A signaling to p53 involves the p19(ARF) tumour suppressor. *Genes Dev.* **12**, 2434–2442 (1998).
14. Stukenberg, P. T. & Kirschner, M. W. Pin1 acts catalytically to promote a conformational change in Cdc25. *Mol. Cell* **7**, 1071–1083 (2001).
15. Fujimori, E., Takahashi, K., Uchida, C. & Uchida, T. Mice lacking Pin1 develop normally, but are defective in entering cell cycle from G(0) arrest. *Biochem. Biophys. Res. Commun.* **265**, 658–663 (1999).
16. Lowe, S. W., Ruley, H. E., Jacks, T. & Housman, D. E. p53-dependent apoptosis modulates the cytotoxicity of anticancer agents. *Cell* **74**, 957–967 (1993).
17. Vogelstein, B., Lane, D. & Levine, A. J. Surfing the p53 network. *Nature* **408**, 307–310 (2000).
18. Bates, S. & Vousden, K. Mechanisms of p53-mediated apoptosis. *Cell Mol. Life Sci.* **55**, 28–37 (1999).
19. Wulf, G. M. *et al.* Pin1 is overexpressed in breast cancer and cooperates with Ras signalling in increasing the transcriptional activity of c-jun towards cyclin D1. *EMBO J.* **20**, 3459–3472 (2001).
20. Ryo, A., Nakamura, M., Wulf, G., Liu, Y.-C. & Lu, K. P. Pin1 regulates turnover and subcellular localization of β-catenin by inhibiting its interaction with APC. *Nature Cell Biol.* **3**, 793–801 (2001).
21. Tetsu, O. & McCormick, F. β-catenin regulates expression of cyclin D1 in colon carcinoma cells. *Nature* **398**, 422–426 (1999).
22. Damalas, A. *et al.* Excess β-catenin promotes accumulation of transcriptionally active p53. *EMBO J.* **18**, 3054–3063 (1999).
23. Kwonseop, K., Pang, K. M., Evans, M. & Hay, E. D. Overexpression of β-catenin induces apoptosis independent of its transactivation function with Lef-1 or the involvement of major G1 cell cycle regulators. *Mol. Biol. Cell* **11**, 3509–3523 (2000).
24. Park, D. S. *et al.* Cyclin-dependent kinases participate in death of neurons evoked by DNA-damaging agents. *J. Cell Biol.* **143**, 457–467 (1998).
25. Buschmann, T. *et al.* Jun NH2-terminal Kinase phosphorylation of p53 on Thr-81 is important for p53 stabilization and transcriptional activities in response to stress. *Mol. Cell Biol.* **21**, 2743–2754 (2001).
26. Gostissa, M. *et al.* Activation of p53 by conjugation to the ubiquitin-like protein SUMO-1. *EMBO J.* **18**, 6462–6471 (1999).
27. Wu, Z. *et al.* Mutation of mouse p53 Ser23 and the response to DNA damage. *Mol. Cell Biol.* **22**, 2441–2449 (2002).
28. Maestro, R. *et al.* Twist is a potential oncogene that inhibits apoptosis. *Genes Dev.* **13**, 2207–2217 (1999).

Supplementary Information accompanies the paper on Nature's website
(<http://www.nature.com/nature>).

Acknowledgements We thank our colleagues at the LNCIB for advice, discussions and critical reading of the manuscript; S. Piazza, F. Agostini and E. Guida for experimental support; M. Oren for suggestions and for providing the luciferase constructs; M. Serrano, B. Amati, R. Maestro, X. Lu, T. Crook and S. Soddu for supplying other reagents; G. Zambetti for advice about the preparation of mouse thymocytes; R. Vidimari and A. Beorchia for helping in γ-irradiation experiments; M. Stebel for production of MEFs and technical assistance; and J. Xiao for discussions and for sharing unpublished data. This work was supported by the Associazione Italiana per la Ricerca sul Cancro (AIRC) and MURST (PRIN Cofin 2000) (G.D.S.). M.G. is an FIRC (Fondazione Italiana per la Ricerca sul Cancro) Fellow.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to G.D.S.
(e-mail: dsal@sci.area.trieste.it).

Histone methylation by the *Drosophila* epigenetic transcriptional regulator Ash1

Christian Beisel*, Axel Imhof†, Jaime Greene*‡, Elisabeth Kremmer§ & Frank Sauer*‡

* Zentrum für Molekulare Biologie der Universität Heidelberg (ZMBH),
Im Neuenheimer Feld 282, 69120 Heidelberg, Germany

† Adolf-Butenandt Institut, Universität München, Schillerstrasse 44,
80336 München, Germany

‡ Department of Biochemistry, 1447 Boyce Hall, University of California,
Riverside, California 92521-0129, USA

§ GSF-Forschungszentrum, Institut für Molekulare Immunologie,
Marchionistrasse 25, 81377 München, Germany

The establishment and maintenance of mitotic and meiotic stable (epigenetic) transcription patterns is fundamental for cell determination and function¹. Epigenetic regulation of transcription is mediated by epigenetic activators and repressors, and may require the establishment, ‘spreading’ and maintenance of epigenetic signals². Although these signals remain unclear, it has been proposed that chromatin structure and consequently post-translational modification of histones may have an important role in epigenetic gene expression^{3,4}. Here we show that the epigenetic activator Ash1 (ref. 5) is a multi-catalytic histone methyl-transferase (HMTase) that methylates lysine residues 4 and 9 in H3 and 20 in H4. Transcriptional activation by Ash1 coincides with methylation of these three lysine residues at the promoter of Ash1 target genes. The methylation pattern placed by Ash1 may serve as a binding surface for a chromatin remodeling complex containing the epigenetic activator Brahma (Brm)⁶, an ATPase, and inhibits the interaction of epigenetic repressors with chromatin. Chromatin immunoprecipitation indicates that epigenetic activation of *Ultrabithorax* transcription in *Drosophila* coincides with trivalent methylation by Ash1 and recruitment of Brm. Thus, histone methylation by Ash1 may provide a specific signal for the establishment of epigenetic, active transcription patterns.

Acetylation/de-acetylation, ubiquitination and methylation of histones (H1, H2A, H2B, H3, H4) have been correlated with the activation and silencing of transcription^{7–10}. Histone methylation occurs predominantly at arginine and lysine residues in the amino-terminal tails of H3 and H4 (ref. 10). Arginine methylation mediates transcriptional activation by hormone receptors and probably other chromatin-dependent processes¹¹. By contrast, methylation of K9 and K4 in H3 and K20 in H4 has been linked to transcriptionally inactive chromatin, and corresponding HMTases have been identified^{10–14}. Methylation of H3 K4 has also been detected in transcription-competent chromatin^{15,16}, but the functional link between histone methylation and activation has not been dissected.

To identify HMTases that establish activation-specific methylation patterns, we used a biochemical screen that identified Ash1, a member of the trithorax group of epigenetic activators^{1,5}, as an HMTase. Ash1 contains a SET domain—the ‘signature motif’ of lysine-specific HMTases—flanked by cysteine-rich regions (pre-SET and post-SET domains)^{5,10}. To confirm that Ash1 has HMTase activity, we tested the ability of recombinant Ash1 derivatives Ash1ΔN (deleted N terminus) and Ash1(SET) (containing the pre-SET, post-SET and SET domains only) to methylate histones (Fig. 1a). The Ash1 derivatives methylated H3 and, to a lesser extent, H4 in polynucleosomes and histone core octamers (Fig. 1a, b). By contrast ‘free’ H3 and H4 were methylated to a lesser extent compared with nucleosomes, even though free histones were present

at a fivefold excess over polynucleosomal histones or when supplemented with DNA (Fig. 1d). These results suggest that Ash1 methylates H3 and H4. As Ash1 used in the described HMTase assays was purified from eukaryotic cells, the HMTase activity of Ash1 could result from an associated rather than intrinsic activity. To test this, we subjected Ash1 Δ N to protein transfer membrane assays that detect intrinsic enzymatic activities in proteins⁸. Ash1(SET) was separated by SDS–polyacrylamide gel electrophoresis (PAGE), transferred electrophoretically onto polyvinylidene fluoride (PVDF) membrane, and denatured/re-natured. Reconstituted Ash1(SET) methylated H3 and H4 (Fig. 1e), suggesting that Ash1 has intrinsic HMTase activity.

To identify the target amino acid residue(s) of Ash1, radiolabelled H3 was subjected to Edman-degradation. Scintillation counting of the released amino acid fractions detected radiolabelling of H3 K4 and K9 (Fig. 2a). To support this, we tested the ability of Ash1 Δ N to methylate peptides consisting of amino acids 1–20 of H3 (H3(1–20)). Ash1 Δ N methylated the peptides H3(1–20), H3(1–20)K4 (which contains H3 K4 but leucine residues instead of lysine residues at positions 9, 14 and 18) and H3(1–20)K9 (which contains H3 K9 but leucine residues instead of lysine residues at positions 4, 14 and 18) (Fig. 2b). By contrast, H3(1–20)L4/L9 peptides, which

contain leucine residues at position 4 and 9 of H3, were not significantly methylated (Fig. 2b), indicating further that Ash1 methylates H3 K4 and K9. Owing to the weak radiolabelling, the target(s) of Ash1 in H4 could not be identified by Edman-degradation (data not shown). As H4 K20 is the only H4 residue being dimethylated *in vivo*, we generated a monoclonal antibody against dimethylated H4 K20 (anti-dim(H4-K20)) to investigate whether Ash1 methylates H4 K20 (refs 10, 14). H4 that was free, in histone core octamers or polynucleosomes, was methylated by Ash1 and analysed by western blot analysis. Anti-dim(H4-K20) antibody recognized Ash1-methylated H4, but not un-methylated H4, indicating that Ash1 methylates H4 K20 (Fig. 2c, d).

Single amino acid point mutations *ash1*¹⁰ and *ash1*²¹ abolish Ash1 activator function in *Drosophila* (ref. 5). The mutation in *ash1*¹⁰ (N1458I) resides within the SET domain, and in *ash1*²¹ (E1357K) in the pre-SET domain (ref. 5). To assess whether these mutations affect HMTase activity, we expressed and purified recombinant proteins containing one of these mutations (Ash1 Δ N¹⁰, Ash1 Δ N²¹) and a third mutant (Ash1 Δ N¹¹⁴²) whose mutation (H1459K) resides in the SET domain and abolishes HMTase activity of SUV39H1 (ref. 12). HMTase assays revealed that the mutants did not significantly methylate H3 and H4, indicating that the mutations abolish HMTase activity (Fig. 2e) and that both the pre-SET and SET domains of Ash1 contribute to HMTase activity and transcriptional activation by Ash1.

To assess whether the mutations in Ash1 specifically inactivate HMTase activity or cause a general functional inactivation, we investigated the ability of mutant Ash1 Δ N to bind the known interaction partner Trx. Ash1 Δ N and the three mutants interacted with Trx *in vitro* (see Supplementary Information), suggesting that the inability of mutant Ash1 proteins to methylate histones is based on a specific inactivation of HMTase activity.

To investigate the effect of *ash1*¹⁰ and *ash1*²¹ on transcriptional activation by Ash1 in *Drosophila*, we used transgenic flies carrying the Ash1-dependent reporter gene *N18/15*, which contains a 4-kilobase (kb) regulatory element of the *bxd* region from the Ash1 target gene *Ubx* fused to the mini-white gene (Fig. 2f)¹⁷. Ash1 supports activation of *N18/15* transcription in the *Drosophila* eye (Fig. 2f). By contrast, *N18/15* expression is significantly reduced in *ash1*^{10/+} or *ash1*^{21/+} heterozygous flies (Fig. 2f). As Ash1²¹ and Ash1¹⁰ lack HMTase activity *in vitro*, these results imply that HMTase activity contributes to transcriptional activation by Ash1 *in vivo*.

To dissect the functional relationship between transcriptional activation and histone methylation by Ash1, we reconstituted transcriptional activation by Ash1 in *Drosophila* S2 cells. To monitor transcription in chromatin, we generated S2 (BCAT5) cells that carry the stable integrated reporter gene *BCAT5*, which contains five DNA-binding sites for the yeast activator Gal4, a core promoter and the bacterial *cat* gene^{18,19}. To recruit Ash1 to chromatin, Ash1 derivatives were fused to the Gal4 DNA-binding domain (amino acids 1–147) (Gal4(DBD))²⁰. BCAT5 cells were transfected with plasmids expressing fusion proteins comprising Gal4(DBD) and either wild type or mutant Ash1 Δ N. Gal4(DBD)–Ash1 Δ N activates *BCAT5* expression 20-fold (Fig. 3a)²⁰, whereas HMTase-inactive Ash1 Δ N derivatives did not (Fig. 3a). These results support the hypothesis that HMTase activity of Ash1 mediates activation of transcription.

To link transcriptional activation by Ash1 to histone methylation, we used crosslinked chromatin immunoprecipitation (XChIP)²¹, which detects protein–DNA interactions *in vivo*. Crosslinked chromatin was isolated from BCAT5 cells expressing Gal4(DBD)–Ash1 Δ N, Gal4(DBD)–Ash1 Δ N¹⁰ or Gal4(DBD)–Ash1 Δ N²¹, and immunoprecipitated by antibodies recognizing dimethylated H3 K4, H3 K9 or H4 K20. Precipitated DNA was purified and the enhancer/promoter of target genes was detected by polymerase chain reaction (ref. 21). All three antibodies precipitated chromatin

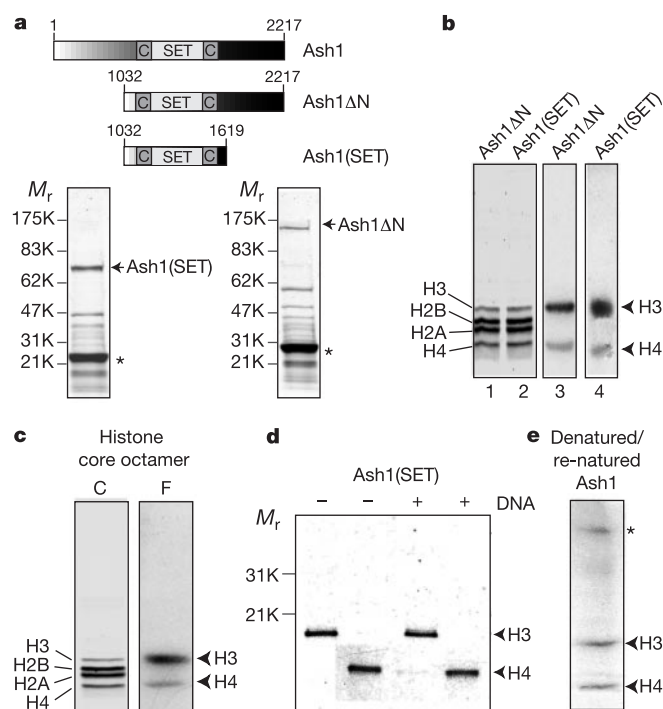


Figure 1 Ash1 methylates histone H3 and H4. **a**, Schematic representation of Ash1 derivatives (top panel). Numbers refer to amino acids in Ash1 (see G117737643; <http://www.ncbi.nlm.nih.gov>). The SET domain (SET) and the cysteine-rich pre-SET and post-SET domains (C) are indicated. The bottom panel shows Coomassie-blue-stained SDS–PAGE gels of immunopurified Ash1(SET) and Ash1 Δ N, whose position is indicated. The asterisk marks the position of the anti-Flag antibody light chain. **b**, Coomassie-blue-stained SDS–PAGE gel (lanes 1, 2) and fluorograms (lanes 3, 4) of HMTase assays programmed with polynucleosomes and Ash1 Δ N (lanes 1, 3) or Ash1(SET) (lanes 2, 4). **c**, Coomassie-blue-stained SDS gel (C) and corresponding fluorogram (F) of an HMTase assay as in **b**, except that reactions contained histone core octamers and Ash1 Δ N. **d**, Fluorogram of HMTase assays programmed with Ash1(SET) and either H3 or H4 in the absence or presence of DNA. **e**, Fluorogram of an HMTase assay programmed with polynucleosomes and protein transfer membrane-bound, denatured/re-natured Ash1(SET). An asterisk marks the position of auto-methylated Ash1(SET). The position of histones (**b–e**) and the position and relative molecular mass (M_r , K) of protein standards is indicated (**a**, **d**).

containing the *BCAT5* enhancer/promoter from cells in which Gal4(DBD)–Ash1ΔN activates transcription (Fig. 3b). Methylation of these lysine residues was detectable 500 bp upstream of the enhancer/promoter and at the 3'-end of the *cat* gene (data not shown). In cells expressing Gal4(DBD)–Ash1ΔN¹⁰, methylation of H3 K4 was undetectable, but weak methylation of H3 K9 and H4 K20 could be observed. This finding supports current models proposing that transcriptional repression correlates with methylation of H3 K9 and H4 K20 (refs 10–14). As, however, H3 K9 and H4 K20 methylation is enhanced at the transcriptionally active (active) reporter, transcriptional activation by Ash1 correlates with *de novo* methylation of not only H3 K4 but also H3 K9 and H4 K20.

Methylation of H3 K9 at the transcriptionally silent (silent) enhancer/promoter implied that *BCAT5* might be associated with HP1, which binds methylated H3 K9. We tested this by XChIP using anti-HP1 polyclonal antibody. The antibody precipitated the *BCAT5* enhancer/promoter from cells expressing HMTase-inactive Gal4(DBD)–Ash1ΔN¹⁰. By contrast, the enhancer/promoter was only weakly precipitated from cells in which Gal4(DBD)–Ash1ΔN activates reporter expression (Fig. 3b). These results suggest that HP1 binds the silent enhancer/promoter and is removed/relocated from the reporter by Ash1-mediated histone methylation.

To investigate whether Ash1 methylates histones to activate

transcription of a natural target gene, we monitored methylation of *Ubx* in *BCAT5* cells. *Ubx* is not expressed in S2-cells²⁴ but PCR with reverse transcription (RT-PCR) indicates that transiently expressed Gal4(DBD)–Ash1ΔN activates expression of this gene in *BCAT5* cells (Fig. 3c). XChIP experiments indicate that H3 K4 is not methylated and that H3 K9 and H4 K20 are only weakly methylated at the silent *Ubx* promoter (Fig. 3b). By contrast, methylation of all three lysine residues is significantly enhanced when Ash1 activates *BCAT5* expression (Fig. 3b), indicating that transcriptional activation of *Ubx* by Ash1 coincides with methylation of H3 K4, K9 and H4 K20.

Genetic data indicate that Ash1 activates *Ubx* expression in imaginal discs of the third leg²⁵. Therefore, to investigate histone methylation by Ash1 in the natural context of the activator, we examined the methylation pattern of *Ubx* in third leg discs by XChIP. We prepared crosslinked chromatin from third leg discs dissected from third instar larvae. Chromatin immunoprecipitations indicated methylation of H3 K4, K9 and H4 K20 at the *Ubx* promoter (Fig. 3d), suggesting that Ash1-mediated methylation of all three lysine residues coincides with epigenetic activation of *Ubx* transcription in *Drosophila*.

On the basis of the result that methylated lysine residues facilitate or inhibit the binding of proteins, we investigated whether the trivalent (H3 K4, K9 and H4 K20) methylation pattern placed by

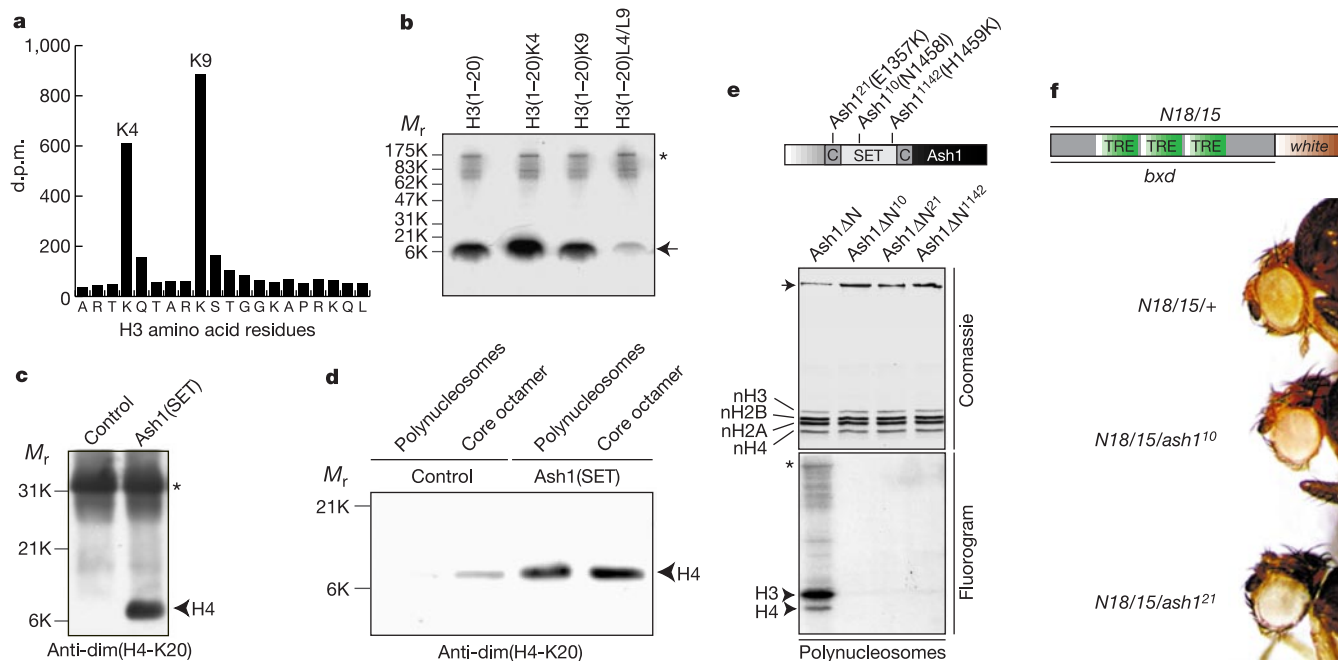


Figure 2 Ash1 methylates K4 and K9 in H3 and K20 in H4. **a**, Microsequencing of radiolabelled polynucleosomal H3. Radiolabelled H3 was subjected to Edman-degradation and the resulting amino acid fractions were analysed by scintillation counting. The x axis shows amino acids 1–20 of H3, the y axis indicates the [³H]-incorporation of individual amino acids in decays per minute (d.p.m.). **b**, HMTase assays using Ash1(SET) and wild-type or mutant H3 peptides (amino acids 1–20). H3(1–20)K4, peptide containing H3 K4 only; H3(1–20)K9, peptide containing H3 K9 only; H3(1–20)L4/L9, peptide containing leucine instead of lysine at positions 4 and 9. The arrow indicates the position of radiolabelled peptide; the asterisk marks auto-methylated Ash1(SET). **c**, Western blot analyses using anti-dimethylated H4 K20 monoclonal antibody (dim(H4-K20)) of HMTase assays containing recombinant H4 and the control protein TAF₁₁₀ or Ash1(SET). The asterisk indicates the position of anti-Flag antibody light chain. **d**, Western blot analyses as in **c**, except that reactions were programmed with polynucleosomes or histone core

octamers, as indicated. The position and relative molecular mass of protein standards is indicated. Note the low basal level of H4 K20 methylation of histone core octamers. **e**, Schematic representation of mutant ASH1ΔN proteins (top). The position and the single amino acid exchange caused by the mutations are indicated. The bottom panel shows a Coomassie-blue-stained SDS gel (top) and fluorogram (bottom) of HMTase assays programmed with the indicated Ash1 derivatives and polynucleosomes. The arrow marks the position of Ash1 derivatives. The asterisk marks auto-methylated Ash1. **f**, Schematic representation of the *N18/15* reporter gene containing the 4-kb regulatory element of the *bxd* region, which contains three Trx response elements (TRE), fused to the *Drosophila* mini-white gene (top). The bottom panel shows photographs of eyes from *N18/15/+* (top), *N18/15/ash1¹⁰* (middle) and *N18/15/ash1²¹* heterozygotes (bottom). Magnification, tenfold.

Ash1 attracts or repels proteins to establish epigenetic activation. The XChIP experiments in Fig. 3b, d indicate that the trivalent methylation pattern removes/relocates HP1 from chromatin. To support this finding, we investigated the interaction of HP1 with methylated H3 peptides and histone core octamers that had been methylated by Ash1 or *Drosophila* SU(VAR)3-9, which methylates H3 K9. HP1 bound H3 K9-methylated peptides and histone core octamers, as well as H3 K4/K9-methylated peptides (Fig. 4a, c). By contrast, HP1 did not bind Ash1-methylated core octamers, suggesting that the trivalent methylation pattern inhibits the interaction of HP1 with chromatin (Fig. 4b).

Protein–protein interaction assays using H3(1–20) peptides methylated at H3 K4, K9 or H3 K4 and K9 (H3 K4/K9), and *Drosophila* embryonic nuclear extract or recombinant proteins, resulted in the identification of three proteins that exhibit differential binding to methylated peptides. Two of these proteins—the epigenetic repressor Polycomb (Pc) and Caf-1 p55, a subunit of different protein complexes involved in, for example, epigenetic repression^{26,27}—bind H3 K9-methylated peptides and histone core octamers, but show significantly reduced binding to H3 K4- or H3 K4/K9-methylated peptides and trivalently methylated histone core octamers (Fig. 4a, b, d–f). Furthermore, protein-binding assays indicate that Brm and Moira (Mor) interact with H3 K4/K9-methylated peptides (Fig. 4g, h). In contrast, both proteins were not recruited to peptides methylated at H3 K4 or H3 K9 (Fig. 4g, h). Brm and Mor are subunits of a SWI/SNF-like chromatin remodelling complex¹, suggesting that this complex, rather than individual proteins, is recruited to K4/K9-methylated H3. These results imply that the trivalent methylation pattern established by Ash1 facilitates or prevents the interaction of proteins with methylated H3 during epigenetic activation. To support this finding, we used XChIP to

investigate the interaction of Brm and repressors with Ash1 target genes. These analyses indicate that Brm and Mor are present at the active but not at silent promoters of Ash1 target genes in cells or third leg imaginal discs (Fig. 3b, d). By contrast, the repressors were only detected at silent promoters (Fig. 3b, d). Thus, transcriptional activation by Ash1 may coincide with the recruitment of Brm and Mor and the extinction of repressor binding at the promoter of Ash1 target genes.

Collectively, our data indicate that the epigenetic activator Ash1 activates transcription by methylation of H3 K4, K9 and H4 K20 at the promoter of target genes. This suggests that epigenetic activation and silencing, which has been linked to methylated H3 K9, may correlate with different histone methylation patterns. Each of the three lysine residues targeted by Ash1 can be individually methylated by specific HMTases, resulting in transcriptional repression and probably activation (H3 K4). Combining these three modifications results in a novel biological readout: epigenetic activation. Why does the trivalent modification pattern generated by Ash1 mediate epigenetic activation? Our results indicate that each modification of the pattern fulfils a specific function. Methylation of H3 K4 prevents the interaction of repressors (Pc, p55) with Ash1 target genes. Methylation of H4 K20 in addition to H3 K4 and H3 K9 prevents the interaction of HP1 with chromatin. Inhibition of repressor binding is an important mechanism, as epigenetic activators and repressors are expressed together during *Drosophila* development. Finally, methylation of H3 K4 and H3 K9 generates an interaction surface for a chromatin-remodelling complex. Our results imply that a specific functional interplay between the epigenetic activators Ash1 and Brm mediates epigenetic activation of transcription. Ash1 initially binds target genes and generates the trivalent histone methylation pattern, which subsequently recruits a

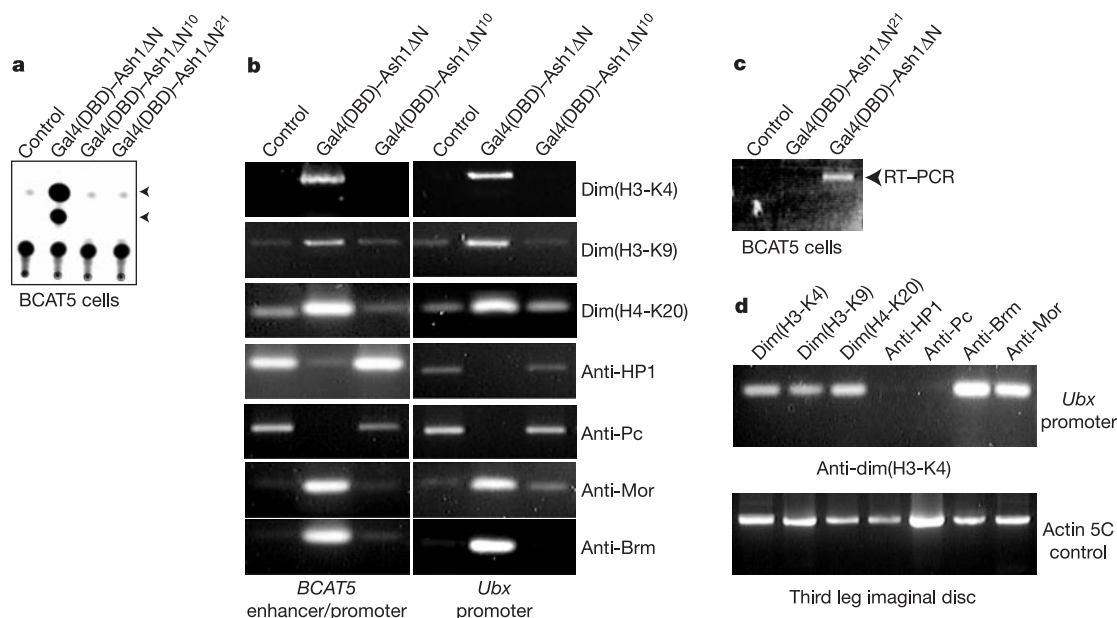


Figure 3 Methylation of K4 and K9 in H3 and K20 in H4 coincides with epigenetic activation of transcription by Ash1. **a**, Autoradiogram of CAT assays using whole-cell extract prepared from BCAT5 cells transfected with plasmids expressing the co-activator TAF_{II}110 (ref. 28) (control, lane 1), Gal4(DBD)-Ash1ΔN (lane 2), Gal4(DBD)-Ash1ΔN¹⁰ (lane 3) or Gal4(DBD)-Ash1ΔN²¹ (lane 4). The arrowheads indicate the position of acetylated [¹⁴C]chloramphenicol. **b**, Photograph of ethidium-bromide-stained agarose gels containing PCR-amplified DNA obtained from XChIP experiments. *In vivo* crosslinked chromatin was isolated from BCAT5 cells that express TAF_{II}110 (control) or the indicated Ash1 derivatives. Crosslinked chromatin was immunoprecipitated using the indicated antibodies. PCR was used to detect a 270-bp fragment of the *BCAT5* enhancer.

c, Ethidium-bromide-stained agarose gel of RT-PCR reactions containing PCR primers generating a 330-bp fragment of the *Ubx* coding region and DNA that was isolated from BCAT5 cells expressing TAF_{II}110 (control) or the indicated Ash1 derivatives. **d**, The top panel shows XChIP experiments as in **b**, but using *in vivo* crosslinked chromatin prepared from imaginal discs of the third leg of *Drosophila*. Shown is the photograph of PCR reactions detecting a *Ubx* promoter fragment in chromatin that had been immunoprecipitated using the indicated antibodies. The bottom panel is a photograph of PCR reactions that detect the presence of the actin 5C promoter in *in vivo* crosslinked chromatin using anti-dim(H3-K4) antibodies (Upstate). Precipitation of the actin 5C promoter was used to standardize PCR reactions shown in the top panel.

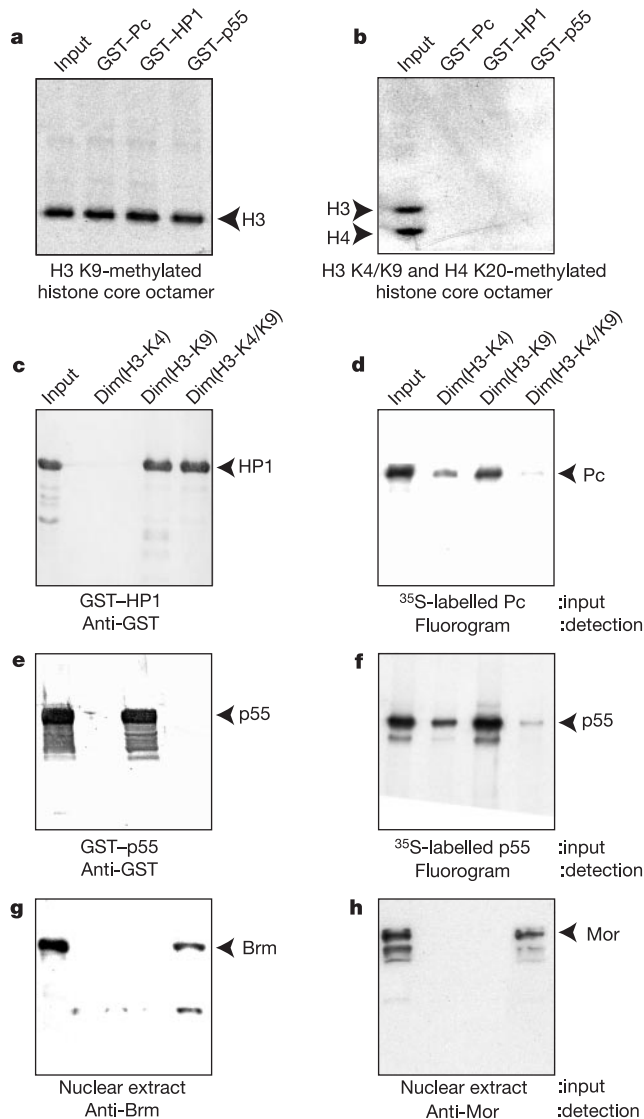


Figure 4 The trivalent methylation pattern generated by Ash1 recruits a Brm-containing chromatin remodelling factor. **a**, Fluorogram of protein–protein interaction assays containing glutathione sepharose loaded with fusion proteins consisting of GST and HP1, Pc or p55, and histone core octamers methylated at H3 K9. **b**, Same as **a**, except that histone core octamers methylated at H3 K4, H3 K9 and H4 K20 by Ash1 were used. **c–h**, Protein–protein interaction assays using H3 peptides (amino acids 1–20) dimethylated at H3 K4 (dim(H3-K4)), H3 K9 (dim(H3-K9)) or both (dim(H3-K4/K9)). **c**, GST–HP1; **d**, *in vitro* expressed ³⁵S-labelled Pc; **e**, GST–p55; **f**, *in vitro* expressed ³⁵S-labelled p55; **g**, **h**, *Drosophila* nuclear extract. Bound proteins were separated by SDS–PAGE and detected as indicated. Input is 10% of the input material. The position of H3 and H4 is indicated.

Brm-containing chromatin-remodelling complex. The activity of this complex may contribute to the establishment of epigenetic active chromatin structures. □

Methods

Plasmids

Expression plasmids were constructed as described in Supplementary Information. Proteins were expressed and purified as described²⁸.

Crosslinked chromatin immunoprecipitation

XChIP was performed essentially as described²¹ with the following modifications. We used either 750 third leg imaginal discs or 4×10^7 BCAT5 cells that were transfected with plasmids expressing wild-type or mutant Gal4(DBD)–Ash1 fusion proteins. At 60 h after transfection, cells were collected and crosslinked with 1.8% formaldehyde. Crosslinked

chromatin was purified as described²¹. Chromatin was immunoprecipitated using anti-dimethylated H3 K4 (dim(H3-K4)), anti-dimethylated H3 K9 (dim(H3-K9)) (both Upstate Biotechnology), and anti-dimethylated H4 K20 (dim(H4-K20)), as well as anti-HP1 anti-Pc, anti-Brm and anti-Mor antibodies. After immunoprecipitation, crosslinked chromatin was incubated with proteinase K (0.5 mg ml^{−1}) at 37 °C for 16 h. Next, crosslink was reversed by incubation of the probe at 65 °C for 6 h. DNA was purified by phenol/chloroform extraction, precipitated with ethanol, and resuspended in 100 μl TE (10 mM Tris (pH 8.0), 1 mM EDTA). For quantitative PCR we used 1/500–1/20 of the precipitated DNA. The presence of a 720-bp fragment from the constitutive active actin 5C promoter, which is methylated at H3 K4, K9 and H4 K20, was detected by PCR to normalize the precipitates obtained from XChIP (C.B. and E.S., unpublished results).

Fly strains

The following fly strains were used: *OregonR* (Bloomington stock centre), *ash1*¹⁰, *ash1*²¹ (ref. 5) and *N18/15/N18/15* that contains the *N18/15* transgene (ref. 17). Flies were maintained under standard conditions. To assess the effect of *ash1*¹⁰ and *ash1*²¹ on the *N18/15* transgene, *N18/15/N18/15* females were crossed with *w¹¹¹⁸; ash1¹⁰/+*, *w¹¹¹⁸; ash1²¹/+* or *w¹¹¹⁸; +/+* males and the eye colour of progenies was analysed. To exclude variations of eye colour resulting from age differences of the flies, we compared the eye colour of 3-day-old progenies expressing wild-type or mutant Ash1. Eyes are shown at a 100-fold magnification and were photographed using a standard dissection binocular.

In vitro histone methyltransferase assay

Methyltransferase assays were performed as described (ref. 16). Ash derivatives were incubated in methyltransferase buffer (25 mM Tris, pH 8.0, 100 mM NaCl, 1 mM phenylmethylsulphonyl fluoride) in the presence of 2 μCi S-adenosyl-L-[methyl-³H]-methionine (50–62 mCi mmol^{−1}) for 1 h at 30 °C. Reactions contained approximately 0.1–0.5 μg wild-type and mutant Ash1 derivatives and 1–3 μg polynucleosomal histones, 5–10 μg recombinant histones or 10 μg histone peptides.

Protein microsequencing

To determine the target amino acids of Ash1 in H3, Ash1-labelled nucleosomal H3 was subjected to microsequencing and fractions containing individual amino acid residues were analysed by scintillation counting as described¹⁶.

In vitro protein-binding assays

Ash1 derivatives, labelled with ³⁵S-labelled methionine, Pc and p55 were generated by *in vitro* expression²⁸. Glutathione S-transferase (GST)–HP1, GST–p55, GST–Pc and GST–Trx (SET) were expressed and purified by affinity chromatography using glutathione-sepharose (Amersham/Pharmacia) according to the manufacturer's instructions²⁸. Protein–protein interaction assays involving Trx were performed as described¹⁷. Binding assays involving H3 peptides were performed as described^{22,23} using *in vitro* expressed proteins or fractionated embryonic *Drosophila* nuclear extract, prepared as described²⁸, as protein source. The identity of bound proteins was detected by western blot analyses as described⁴.

Drosophila tissue culture

Drosophila S2 Schneider cells were maintained and transfected with DNA as described²⁰, and cells containing a stable *BCAT5* reporter gene were generated¹⁹. CAT assays were performed as described²⁰.

Preparation of polynucleosomes

Polynucleosomes were prepared from nuclei essentially as described²⁸ with the following modifications: after nuclease digestion, the chromatin-containing supernatant was dialysed against TE-600 (10 mM Tris, pH 8.0, 0.1 mM EDTA, 600 mM NaCl) and further purified by gel filtration using a Superdex-200 column. Polynucleosomes used in methyltransferase assays contained on average 2–20 nucleosomes.

Received 10 July; accepted 16 September 2002; doi:10.1038/nature01126.

Published online 9 October 2002.

- Simon, J. A. & Tamkun, J. W. Programming off and on states in chromatin: mechanisms of Polycomb and trithorax group complexes. *Curr. Opin. Genet. Dev.* **12**, 210–218 (2002).
- Urnov, F. D. & Wolffe, A. P. Above and within the genome: epigenetics past and present. *J. Mammary Gland Biol. Neoplasia* **6**, 153–167 (2001).
- Turner, B. M. Histone acetylation and an epigenetic code. *Bioessays* **22**, 836–845 (2000).
- Cavalli, G. & Paro, R. Epigenetic inheritance of active chromatin after removal of the main transactivator. *Science* **286**, 955–958 (1999).
- Tripoulas, N. A., Hersperger, E., La Jeunesse, D. & Shearn, A. Molecular genetic analysis of the *Drosophila melanogaster* gene absent, small or homeotic discs1 (*ash1*). *Genetics* **137**, 1027–1038 (1994).
- Elfring, L. K. *et al.* Genetic analysis of brahma: the *Drosophila* homolog of the yeast chromatin remodeling factor SWI2/SNF2. *Genetics* **148**, 251–265 (1998).
- Kuo, M. H. & Allis, C. D. Roles of histone acetyltransferases and deacetylases in gene regulation. *Bioessays* **20**, 615–626 (1998).
- Pham, A.-D. & Sauer, E. Ubiquitin-activating/conjugating activity of TAF_{II}250, a mediator of activation of gene expression in *Drosophila*. *Science* **289**, 2357–2360 (2000).
- Sun, Z. W. & Allis, C. D. Ubiquitination of histone H2B regulates H3 methylation and gene silencing in yeast. *Nature* **418**, 104–108 (2002).
- Zhang, Y. & Reinberg, D. Transcription regulation by histone methylation: interplay between different covalent modifications of the core histone tails. *Genes Dev.* **15**, 2343–2360 (2001).

11. Stallcup, M. R. Role of protein methylation in chromatin remodeling and transcriptional regulation. *Oncogene* **20**, 3014–3020 (2001).
12. Rea, S. *et al.* Regulation of chromatin structure by site-specific histone H3 methyltransferases. *Nature* **406**, 593–599 (2000).
13. Briggs, S. D. *et al.* Histone H3 lysine 4 methylation is mediated by Set1 and required for cell growth and rDNA silencing in *Saccharomyces cerevisiae*. *Genes Dev.* **15**, 3286–3295 (2001).
14. Nishioka, K. *et al.* PR-Set7 is a nucleosome-specific methyltransferase that modifies lysine 20 of histone H4 and is associated with silent chromatin. *Mol. Cell* **9**, 1201–1213 (2002).
15. Nishioka, K. *et al.* Set9, a novel histone H3 methyltransferase that facilitates transcription by precluding histone tail modifications required for heterochromatin formation. *Genes Dev.* **16**, 479–489 (2002).
16. Strahl, B. D., Ohba, R., Cook, R. G. & Allis, C. D. Methylation of histone H3 at lysine 4 is highly conserved and correlates with transcriptionally active nuclei in *Tetrahymena*. *Proc. Natl Acad. Sci. USA* **96**, 14967–14972 (1999).
17. Rozovskaia, T. *et al.* Trithorax and ASH1 interact directly and associate with the trithorax group-responsive *bxd* region of the Ultrathorax promoter. *Mol. Cell. Biol.* **19**, 6441–6447 (1999).
18. Lillie, J. W. & Green, M. R. Transcription activation by the adenovirus E1a protein. *Nature* **338**, 39–44 (1989).
19. Rio, D. C. & Rubin, G. M. Transformation of cultured *Drosophila melanogaster* cells with a dominant selectable marker. *Mol. Cell. Biol.* **5**, 1833–1838 (1985).
20. Sauer, F. & Jäckle, H. Dimerization and the control of transcription by Krüppel. *Nature* **364**, 454–457 (1993).
21. Kuo, M. H. & Allis, C. D. *In vivo* cross-linking and immunoprecipitation for studying dynamic Protein:DNA associations in a chromatin environment. *Methods* **19**, 425–433 (1999).
22. Bannister, A. J. *et al.* Selective recognition of methylated lysine 9 on histone H3 by the HP1 chromo domain. *Nature* **410**, 120–124 (2001).
23. Lachner, M., O'Carroll, D., Rea, S., Mechtler, K. & Jenuwein, T. Methylation of histone H3 lysine 9 creates a binding site for HP1 proteins. *Nature* **410**, 116–120 (2001).
24. Orlando, V. & Paro, R. Mapping Polycomb-repressed domains in the *bithorax complex* using *in vivo* formaldehyde cross-linked chromatin. *Cell* **75**, 1187–1198 (1993).
25. LaJuenesse, D. & Shearn, A. Trans-regulation of thoracic homeotic selector genes of the Antennapedia and bithorax complexes by the trithorax group genes: absent, small, and homeotic discs 1 and 2. *Mech. Dev.* **53**, 123–139 (1995).
26. Paro, R. & Hogness, D. S. The Polycomb protein shares a homologous domain with a heterochromatin-associated protein of *Drosophila*. *Proc. Natl Acad. Sci. USA* **88**, 263–267 (1991).
27. Ridgway, P. & Almouzni, G. CAF-1 and the inheritance of chromatin states: at the crossroads of DNA replication and repair. *J. Cell Sci.* **113**, 2647–2658 (2000).
28. Sauer, F., Hansen, S. K. & Tjian, R. DNA template and activator-coactivator requirements for transcriptional synergism by *Drosophila* Bicoid. *Science* **270**, 1825–1828 (1995).

Supplementary Information accompanies the paper on *Nature's* website (<http://www.nature.com/nature>).

Acknowledgements We thank M. Meisterernst for anti-dimethylated H4-K20 antibody; A. Shearn and A. Mazo for fly strains; R. Deuring and J. Tamkun for anti-Brm antibody; S. Schönfelder and R. Paro for Trx cDNA; J. Tyler for p55 cDNA; and various members of the laboratory for critical reading of the manuscript. We are grateful to J. Rami for editorial help and Y. Cully for photowork. This work was supported by a Graduiertenkolleg 484 fellowship (Deutsche Forschungsgemeinschaft) (C.B.) and a Tranregio/Sonderforschungsbereich grant (E.S.).

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to F.S. (e-mail: f.sauer@mail.zmbh.uni-heidelberg.de).

erratum

Self-organization of supramolecular helical dendrimers into complex electronic materials

V. Percec, M. Glodde, T. K. Bera, Y. Miura, I. Shiyonovskaya, K. D. Singer, V. S. K. Balagurusamy, P. A. Heiney, I. Schnell, A. Rapp, H.-W. Spiess S. D. Hudson & H. Duan

Nature **419**, 384–387 (2002).

In this Letter, the volume numbers appeared incorrectly in the footer of pages 384, 385 and 386 as 'VOL 417'. It should have read 'VOL 419'. □

Contacts

Publisher: Ben Crowe

Editor: Paul Smaglik

Marketing Manager: David Bowen

European Head Office, London

The Macmillan Building
4 Crinan Street
London N1 9XW, UK
Tel +44 (0) 20 7843 4961
Fax +44 (0) 20 7843 4996
e-mail: naturejobs@nature.com

Senior European Sales Manager:
Nevin Bayoumi (4978)

UK/ RoW/ Ireland:

Matt Powell (4953)

Andy Douglas (4975)

Frank Phelan (4944)

Netherlands/ Italy/ Iberia:

Evelina Rubio Hakansson (4973)

Scandinavia: Silje Opstrup (4994)

France/ Belgium:

Amelie Pequignot (4974)

Production Manager:

 Billie Franklin

To send materials use London
address above.

Tel +44 (0) 20 7843 4814

Fax +44 (0) 20 7843 4996

e-mail: naturejobs@nature.com

International

Advertising Coordinator:

Hind Berrada (4935)

Naturejobs web development:

Tom Hancock

Naturejobs online production:

Ben Lund

European Satellite Office

Germany/ Austria/ Switzerland:

Patrick Phelan, Odo Wulffen

Tel + 49 89 54 90 57 11/-2

Fax + 49 89 54 90 57 20

e-mail: p.phelan@nature.com

o.wulffen@nature.com

US Head Office, New York

345 Park Avenue South,
10th Floor, New York, NY 10010-1707

Tel +1 800 989 7718

Fax +1 800 989 7103

e-mail: naturejobs@natureny.com

US Sales Manager:

 Peter Bless

US Advertising Coordinator:

Linda Adam

Japan Head Office, Tokyo

MG Ichigaya Building (5F),

19-1 Haraikatamachi,

Shinjuku-ku,

Tokyo 162-0841

Tel +81 3 3267 8751

Fax +81 3 3267 8746

e-mail: k.johnson@naturejpn.com

Asia-Pacific Advertising Manager:

Kevyn Johnson

naturejobs

Grappling for grants

Times are changing for young biomedical researchers in the United States. Today, youngsters either take longer to secure their first tenure-track post or fail to compete effectively for grants with their more established colleagues.

At least, that is the picture painted by data released recently by the National Institutes of Health (NIH). Covering who gets what sort of funding, the data reveal that in 1980 grants to investigators under the age of 35 accounted for 23% of all funding awards, but by last year that figure had dropped to just 4%.

Academics who wish to blame this change on the NIH can point to the agency's increased focus on few large projects rather than myriad small ones, and a shift away from training grants. And NIH officials who want to indict academia could argue that universities watching their bottom lines are creating fewer permanent positions, and so are increasingly relying on adjuncts and postdocs to handle teaching and research.

The truth? Probably a combination of both arguments, plus a few other factors. Issues such as demographics must surely be important. There is evidence that the upper end of the academic pipeline is becoming clogged with senior scientists. And although younger investigators are finding it hard to win funds, the actual pool of cash available has grown dramatically, effectively doubling in the past five years alone.

So what would help young investigators in their first academic post? One thing the NIH could do is reinstate training grants — but without some of the baggage attached to earlier, now abolished, programmes, such as the need for reference letters or stretching relatively small grants over a long period of time. In the meantime, young scientists must continue to compete with their more established peers, and should treat the grant-seeking process like an election — they should apply early and often.

Paul Smaglik

Naturejobs editor



Contents

WWW.NATUREJOBS.COM

Career centre
Information on the
scientific job market

FOCUS

SPOTLIGHT

RECRUITMENT

SCIENTIFIC ANNOUNCEMENTS

SCIENTIFIC EVENTS